The Thesis Committee for Wonseuk Son
Certifies that this is the approved version of the following thesis:

Exploring the Feasibility of Measuring Individual Labor Productivity Using a Wearable Activity Tracker

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________
Carlos H. Caldas

Co-supervisor: ______________________________
John D. Borcherding

Exploring the Feasibility of Measuring Individual Labor Productivity Using a Wearable Activity Tracker

by

Wonseuk Son

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

August, 2017

Dedication

To my great parents, Soon-Mo Son and Hee-Soon Kim.

To my supportive brother, Won-Il Son.

Acknowledgements

I would like to thank all those who have helped me along the way in carrying out this thesis. First, I appreciate my advisor, Dr. Caldas for his patience, support and encouragement. I could not properly complete this thesis without his guidance and precious advising. His sharp insight and deep knowledge have provided various points of view for me to make this research more robust. In addition, activity analysis that I learned from his class, Front End Planning, was a good guidance to me in establishing the framework of this new methodology.

Next, I want to express my appreciation to my second reader, Dr. Borcherding, for his patience and detailed reviews. Especially, his Productivity class was one of the biggest motivations for me to choose productivity as a main topic of this thesis. The class materials regarding productivity I went through in the class helped me conduct this research from various perspectives.

Moreover, it would have been more hard time to complete the thesis without my fellow students. First, I would like to thank Jeyoung Woo for his constant support, encouragement, and guidance. He has been always ready to help me or others like an angel. I will not forget his precious help. I also want to say my gratitude to Bongyoung Yoo and Euijin Yang. They always made me smile and laugh while conducting this thesis during the whole semesters.

Lastly, I have been owing my parents and brother an astronomical debt of gratitude for their infinite love and support. They have been always by my side everywhere and every time. It is the most precious energy that always motivates me.

Abstract

Exploring the Feasibility of Measuring Individual Labor Productivity Using a Wearable Activity Tracker

Wonseuk Son, M.S.E.

The University of Texas at Austin, 2017

Supervisor: Carlos H. Caldas

Productivity measurement in the construction industry has received enormous attention from industry and academia. One of the most crucial factors for achieving high performance in a construction project is labor management. Despite technological advances, construction projects are still driven by labor intensive work. Onsite work productivity determines, to a great extent, the performance of construction projects. However, determining accurate productivity metrics remains a challenge for project managers.

Productivity measurements are still collected or recorded manually from construction projects. Productivity assessments require that well-trained staff perform the measurements, and that considerable time is spent accurately collecting and analyzing the data. Moreover, whether the common data used for productivity calculations are accurate is questionable due to undocumented/unrecorded data.

Productivity is measured at various levels such as by industry, by project, and by activity. Currently, there is no reliable way to measure productivity at the individual laborer

level. Such measurement could provide detailed and accurate information about project productivity, if data from an appropriate number of workers is collected. Identification of the poor productivity performance would be improved using productivity recorded per second with personal and location data instead of hourly, daily or weekly productivity summaries.

The principal objective of this research is to explore the potential of a new productivity measurement methodology that automatically collects laborer data used for calculating productivity. The basic concept is to utilize a wrist activity tracker for data collection process. It mainly consists of automated data collection, and automated/semi-automated data analysis. The basic work plan for extracting direct work hours from total work hours is to identify, understand, and input specific patterns that appear in each of the different activities in the data analysis process.

For proof of concept, field experiments were conducted at three different sites. Results from actual observations (ground truth), and automated/semi-automated data analysis per experiment were compared to evaluate the proposed method.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Although it is difficult to prove that productivity measurement has a direct effect on construction performance improvement, regular productivity measurements during construction allow owners and contractors to identify and remove factors that could delay projects (Park 2005).

Productivity measurement in the construction industry has received attention from industry and academia. However, in spite of the importance of productivity measurement, there are too many factors impacting productivity to accurately quantify the rate of improvement. Productivity rates are calculated and defined in several ways according to the characteristics of work, projects, laborers and the reason for measuring productivity. A typical example of a productivity calculation is the ratio of output to input, such as installed material to planned installation of material or expended cost to hours worked. However, determining accurate productivity measurement techniques remains a challenge for project managers.

1.1. RESEARCH MOTIVATION

Construction productivity is measured at various levels, such as by industry, by project, and by activity (Thomas et al 1990). However, despite its importance there is no reliable way to measure productivity at the individual laborer level. Most construction projects are completed by a crew, a group of laborers, so it is difficult to objectively measure the productivity of an individual laborer in the crew.

Productivity at the individual laborer level can provide sufficiently detailed and accurate information about a construction project, if an adequate number of workers is observed. When management examines labor productivity data, it is difficult to determine which crew or laborer is accountable or when and where poor productivity happened. Identification of the poor productivity could be improved using personal productivity that is recorded per second with physical and location data instead of hourly, daily or weekly productivity summaries.

Productivity measurements in construction projects are implemented with various types of data. Most data are collected or recorded manually from construction projects with paper or electronic documentation. Productivity measurements require that well-trained staff perform the measurements, and that considerable time is spent accurately collecting and analyzing the data.

Moreover, it is questionable whether the common data used for productivity calculation such as workhours, actual cost, and quantity of materials are fully accurate. Undocumented/unrecorded data can occur due to human error. Workhours from additional unplanned crews that subcontractors unofficially assign in order to make up delays may not be recorded or reported to contractors. Although the official record of contractors may indicate that productivity of the work has an average rate and the project is on time and on budget from the contractors' perspective, the actual productivity with undocumented workhours will indirectly impact the planned schedule. Therefore, data collection and analysis should be automated and simplified as much as possible in order to reduce human error in productivity measurement.

1.2. RESEARCH OBJECTIVES

The principal objective of this research is to explore the feasibility of a new personal productivity measurement methodology that can automatically collect and analyze individual data that can be used for calculating productivity. Personal productivity measurement in this research is defined as quantification of the time spent by an individual craft worker on productive and non-productivity activities, calculated as a percentage of time that the worker spent on direct work and other activity categories.

The sub-objectives are as follows:

1. Understand specific patterns of data collected from a wrist activity tracker.
2. Develop a semi-automated data analysis for automatic assignment of activity categories.
3. Demonstrate the feasibility of collecting personal productivity data.

1.3. RESEARCH SCOPE

Experiments for demonstrating the personal productivity measurement methodology were conducted by collecting and analyzing data from on a single laborer per project. Productivity measurements at the activity and project levels were not conducted. Thus, this thesis focuses on the feasibility of measuring personal productivity of each individual laborer, not productivity of the activity or the whole project.

In order to record clear patterns from the data collected on task performance, the type of work chosen for collection should be labor intensive and involve physical activity. In addition, this collection method employs a global positioning system (GPS) to detect the movement of laborers, so the measurement effort focused on outdoor labor. Considering the above conditions, the author chose to collect data from carpenters working on formwork outside who joined in field tests for initial sample data.

The formula for productivity measurement in this research is the ratio of direct work hours to total work hours of a laborer (Eq. 1).

$$\text{Productivity rate} = \frac{Direct\ work\ hours}{Total\ work\ hours} \qquad \text{(Eq. 1)}$$

1.4. THESIS STRUCTURE

Following the Chapter 1 introduction, Chapter 2 reviews previous research in terms of different methodologies for productivity measurement, definition of productivity, and activity tracking. In Chapter 3, the personal productivity measurement is introduced. It includes data collection, data analysis, and demonstration of the method. Regarding the data analysis, two different methodologies are implemented in order to assess the results. Chapter 4 provides results and discussion in terms of the proposed method. Finally, Chapter 5 draws conclusions that include a review of the research, limitations of this research, contributions, and recommendations.

Chapter 2: Background Review

In this chapter, the literature and basic knowledge related to this research is introduced and reviewed.

2.1 Construction Productivity

Construction productivity is an important variable that many researchers and practitioners have studied because productivity directly and indirectly relates to project performance (CII 2006). Construction project managers monitor the current status of the project and seek to improve a project's productivity by measuring its productivity against an industry productivity standard. Two goals of productivity measurement are to keep projects on budget and schedule and to acquire data that can be used for estimating and comparing future projects (Thomas and Mathews 1986).

Construction productivity has been defined in many different ways. Additionally, a variety of methods, each with different formulas, are used to measure productivity in the construction industry (Benzekri 2010). Therefore, it is difficult to compare the productivity of different construction projects. The complexity and the unique features of construction projects make it difficult to determine a standard methodology for construction productivity measurement (Oglesby et al. 1989). Many researchers have tried to define the most effective description for productivity in the construction industry (Oglesby et al. 2002), and the process continues to evolve.

The basic concept for productivity calculation that productivity measurements are based on is the ratio of output to input or input to output (Park 2002). Types of input and output are different according to the methods of productivity measurement, such as work hours for input in unit rate productivity and total cost expended for input in the multifactor productivity of Bureau of Labor Statistics (BLS).

2.1.1. Construction Productivity Measurement

At the various levels of industry, corporate, and project level, the construction productivity measurement method should be different.

For industry level economic models, researchers have developed economic growth accounting equations (Solow 1961; Domar et al. 1961; Jorgenson et al. 1967; Jorgerson et al; 1987). The traditional equations are used in formulating the growth of multifactor productivity for nonmanufacturing industries, measured as output per unit of labor, capital, and other measurable inputs (Harper at el. 2010). One equation is as follows:

$$\frac{\dot{A}}{A} = \frac{\dot{Q}}{Q} - \left(w_k \frac{\dot{K}}{K} + w_l \frac{\dot{L}}{L} + w_e \frac{\dot{E}}{E} + w_m \frac{\dot{M}}{M} + w_s \frac{\dot{S}}{S}\right) \qquad \text{(Eq. 2)}$$

Where a dot over a variable denotes a derivative regarding time and the variables and weights used for factors in these equations are as follows:

A = multifactor productivity,
Q = sectoral output,
K = capital input,
L = labor input,
E = energy input,
M = materials input,
S = purchased business services input,
w_k= weight for capital,
w_l = weight for labor,
w_e = weight for energy,
w_m = weight for materials, and
w_s = weight for business services.

However, the equation of multifactor productivity is not appropriate for the individual project level productivity measurement since the inputs of the equation are difficult to quantify and collect (Kim 2015).

Factor Productivity is more suitable with accuracy and input availability for individual construction projects (Thomas et al. 1990). It is calculated using the ratio of physical output (units) to monetary values of labor, material, and equipment as shown in the Eq. 3. The output is a representative unit of a different types of construction projects: miles for a highway project, rooms for a hotel, square feet for a commercial building, and capacity for a power plant (Kim 2015).

$$Factor\ Productivity = \frac{Physical\ Output\ (Units)}{Total\ cost\ of\ (labor+material+Equipment)} \qquad \text{(Eq. 3)}$$

The factor productivity is often used for the feasibility analysis of a project, but it is not sufficiently accurate and detailed for management to assess project performance.

Unit rates is an activity level productivity measurement that is defined as the ratio of workhours (actual work hours) to quantity (installed quantity). It is widely used in construction research (Thomas et al 1987; Sonmez et al 1998). Construction companies use unit rates frequently as an internal data source for their competitive price in bidding or for assessing subcontractors (Kim 2015).

2.1.2. Construction Productivity Field Data Collection

Field data collection in the construction industry is conducted by on-site observations. Work sampling is the one of the most frequently used methods for construction productivity assessment in both academia and industry. It is a method that calculates and assesses the proportion of time spent on direct work, essential contributory work, and ineffective work of laborers by a series of on-site observations (Thomas and Daily 1983). It was introduced to the construction industry in the 1960s. Its definition and application were elaborated by Thomas and Daily (1980, 1981). Work sampling was verified as a productivity indicator that has a close correlation with labor productivity (Liou and Borcherding 1986). This method is not sufficient to identify details of causes leading to low productivity, so crew-level work sampling supplemented by foremen-delay surveys and craft questionnaires were developed (Tsehayae and Fayek 2012).

Activity analysis is a comprehensive process that is extrapolated from work sampling (Gouett et al. 2011). The first result of work sampling on a project can provide baseline productivity for later cycles of activity analysis (Kim 2015). Activity analysis consists of 5 cyclical steps, as shown in Figure 1. In the Plan Study step, objectives, scope, sample size, activity categories, and observers are determined. Then, the observer starts the Sample step. With an understanding of the results of activity rates, issues that negatively affects the project are identified with other assessment methods, such as foreman-delay surveys and questionnaires in the Analysis step. Then, in the Plan Improvement step, managers plan alternatives or solutions for the issues by focusing on root causes detected from the result of the previous steps. The last step, Implement Improvements, is where the actual implementation of productivity improvement initiatives takes place.

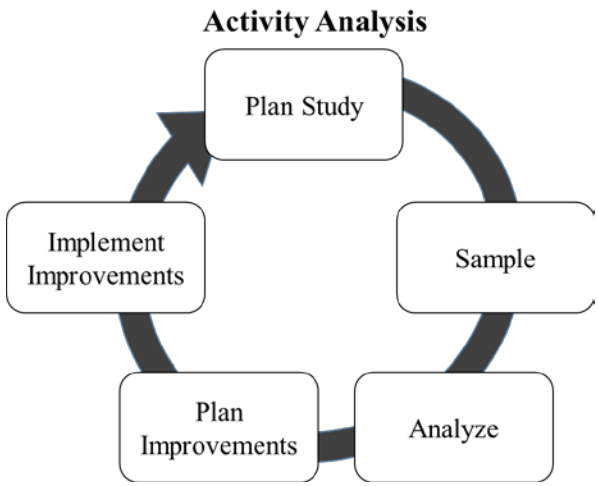


Figure 1. Process of activity analysis (Gouett et al. 2011)

2.2. DATA ANALYSIS FOR PHYSICAL ACTIVITY

In sports science academia, the data of one's heart rate and movement has been common and leading research area for physical activity analysis. Especially, heart rate response is one of the best source to identify physical characteristics of sports. Loftin et al. (1996) used a Polar Vantage heart watch to measure heart rate of male handball players playing handball. Heart rate measured during handball match was averaged and ranged according to playing time in order to find when the players play hard or loose. Barbero-Alvarez et al. (2008) collected and analyzed heart rate, speed, and distance of futsal, 5 vs 5 soccer, players to examine physiological loads during futsal competition and patterns of the collected data. The patterns of heart rate and movement data were detected and categorized according to six activities: "Standing", "Walking", "Jogging", "Medium-intensity running", "High-intensity running", and "Sprint". In 2010, Castellano et al. found patterns of heart rate and GPS data of five beach soccer players according to six different activities and calculate the percentage of each activity and *work:rest* ratio. Neural networks to classify types of activity among children were developed and evaluated (Galindo-Garre et al. 2012)

Fjørtoft et al. (2009) tracked movement patterns and physical activity of children outside using GPS and heart rate monitoring in order to plan landscaping of two schoolyards. Gatti et al. (2011) conducted preliminary validation of three wearable physiological status monitors that measures and evaluates worker's physical strain for safety and health. They assessed three wearable physiological status monitors that measures heart rates and acceleration for measuring and evaluating worker's physical strain. The authors could not find specific correlation coefficients between heart rates and each of activities, "Static", "Thoracic Rotation", "Arm lift", "Batting", "Weight moving", and "Walk on a treadmill", but they illustrated descriptive statistics of correlation between acceleration and the activities. Hsiao et al. (2012) collected and analyzed data about heart rate and oxygen consumption to distinguish different types of work activity under a lab setting. It provided meaningful results that slightly different patterns were showing in different activities, but the patterns were not clear enough to classify the activities because the research focused on only average values of collected data.

In the construction research, few research studies regarding activity tracking were conducted.

2.3. CONCLUSION

Many definitions and measurement methods of construction productivity have been developed in different perspectives at different levels. However, personal productivity measurement methods at the laborer level has not been studied properly, although it can provide detailed information about projects from laborer's perspective. In addition, the manual process of construction productivity assessment still exists in data collection and data analysis plus it is time consuming.

In the science sports domain, there were many studies using physical data such as heart rate and body movement for studying physical activity and they showed good results. Some researchers in the construction academia have also tried to study physical conditions for laborer's safety, health and activity classification.

A few studies were conducted using physical data. However, few research studies, for productivity measurement using personal physical data have been studied in the construction domain Hsiao et al. (2012). The study were conducted under a lab setting with large equipment that restricted laborer's motion and cannot be used onsite. It utilized minute-based data that might not capture enough the changes between activities. Therefore, this thesis will focus on measuring personal productivity onsite by automated data collection and automated/semi-automated data analysis using data recorded per second.

Chapter 3: Semi-Automated Personal Productivity Measurement

The basic concept of the new personal productivity measurement is to utilize a wrist activity tracker for the data collection process. In the traditional activity analysis, the data collection is conducted by an observer walking around the construction site and recording any activity that was observed at regular intervals of time. Observations of the activity analysis were collected from almost all laborers working at the construction site. In contrast, personal productivity measurement collected data used for productivity calculations from an individual laborer wearing a wrist activity tracker. Then, using the collected data, the author classified each second of a worker's activity into 4 different activity categories: Direct Work, Idling, Moving, and Moving (High HR) by analyzing data combinations of GPS data, heart rates, speed, and steps which are collected and recorded in the wrist device. The details of the types of activity are described in Table 1.

	Description	Example
Direct work	"The act of either exerting physical effort to perform a (construction) activity or physically assisting in these activities" (CII 2010)	Hammering, Handling Materials, Cutting, Welding, etc.
Idling	Any aimless activity that is not related to direct work	Chatting, Waiting, Resting, etc.
Moving	Unsupportive movements not related to direct work	Looking for Tools, Traveling, Loitering, etc.
Moving (High HR)	Supportive movements related to direct work	Transporting, Slight supportive movements during direct work, etc.

Table 1. Detailed information about activity types.

Before starting a field test of the new methodology, the author designed the study to employ patterns that are used for classifying data according to the four types of work activity (Table 2). The static movements recorded in the table are speed, steps, and GPS data collected from a laborer that does not provide a specific number or shape that can be regarded as little movement. In contrast, the active movements recorded indicate that at least two data points out of speed, steps, and GPS data show specific numbers.

Expected Basic Patterns	**Possible status**
High heart rates with static movements	Direct Work
Low heart rates with static movements	Idling
Low heart rates with active movements	Moving
High Heart rates with active movements	Moving (High HR)

Table 2. Expected possible patterns.

Wrist Activity Tracker

The activity tracker used in this research for productivity measurement is the Garmin Vivoactive HR (Figure 2). It is worn on the wrist of a worker and records GPS data, heart rates, and steps. Heart rate is measured per second by optical heart rate sensors built into the device. In addition to the heart rate sensor, a GPS sensor, accelerometer, compass, and altimeter are included in the device. The data collected from the worker is automatically synchronized with a smartphone using Bluetooth technology and transmitted to a web-based database shown in Figure 3 for users to check their status at any time through internet access. The data can be also exported in various file extensions for use in various software applications.

Figure 2. Garmin Vivoactive HR (Images adapted from DC Rainmaker).

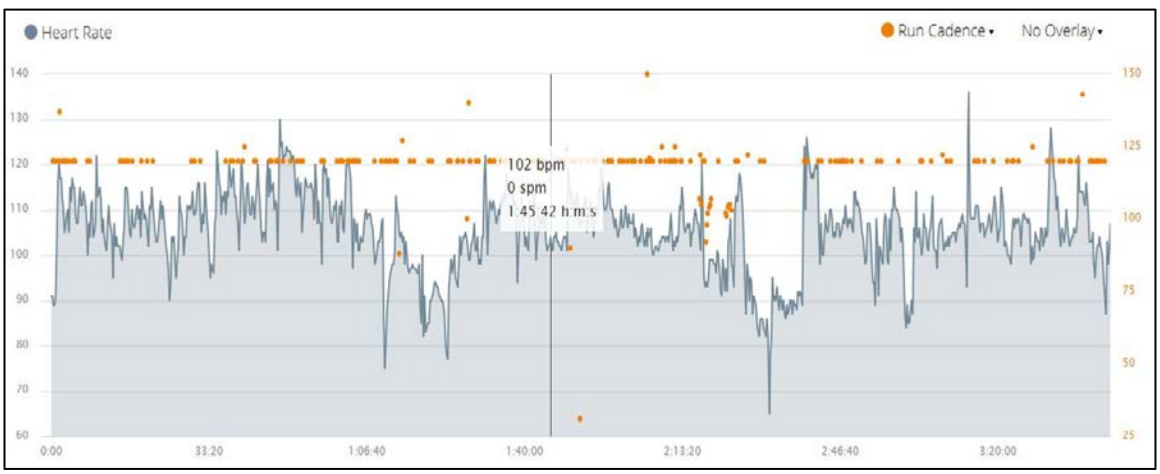


Figure 3. Example of data stored in a web-based database.

Heart Rate

In the measurement, heart rate data is used to separate intense activities from mild activities. The heart rate of workers increases in correlation to use of muscle tissue; the more muscles are used during physical activity, the more oxygen is biologically required, which allows the heart rate variance to differentiate between direct work and idling. The personal spectrum of heart rate, however, differs according to the worker's age, body condition, job, weather condition, and altitude. Therefore, heart rate criteria that is used to classify physical condition is individually defined and adjusted from worker to worker. Figure 4 shows data captured from a small data sampling of the author's experimental usage that illustrates how the heart rates can be zoned according to physical activity. When performing normal work or intensive work, heart rates are clearly shown to be higher than the relaxing zone.

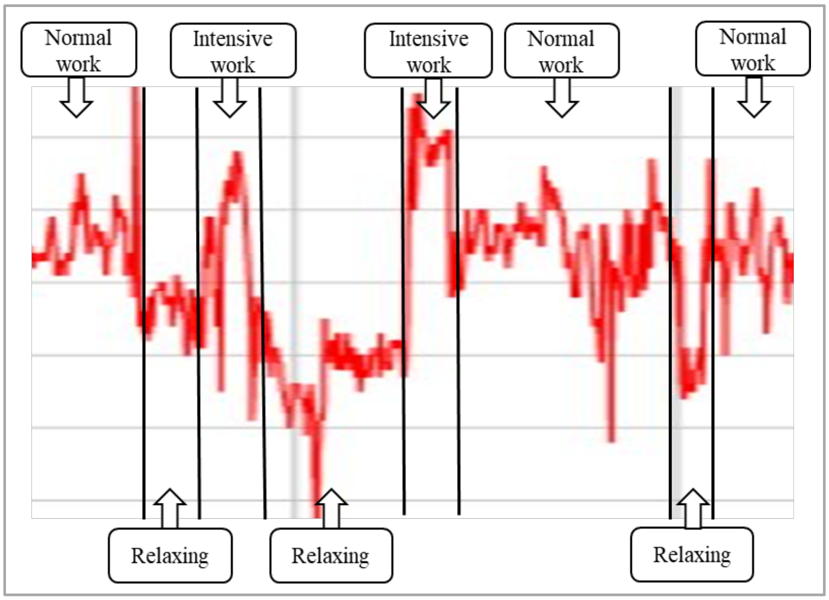


Figure 4. Example of heart rate zoning.

Global Positioning System

GPS data, used to record the distance that a worker moved per second, is analyzed to determinate a change in activity state. GPS coordinates recorded in the form of latitude, longitude, and altitude are illustrated as track points on Google Earth **(**Figure 5**)**, which are used as reference points to detected activity patterns by comparing the points with other numerical data.

Figure 5. Track points on the google earth.

Accelerator

The accelerator tracker in the wrist device provides information on speed and steps. It is not reliable to use accelerator data in isolation when classifying a worker's movements because rapid swings or movements of the worker's wrist can generate misleading data. The accelerator data provides supplementary information to support GPS data recorded at the same time.

3.1. Semi-Automated Personal Productivity Measurement Framework

The productivity measurement methodology has two unique processes (Figure 6). The first process is automated data collection. In the data collection process, a wrist activity tracker worn by a laborer in the field collects heart rates, steps, and GPS data. Then, the data is transferred to a smartphone by Bluetooth and transmitted to a web-based database for storage. The data can be exported in various file formats from the database. The author exported the data in a comma separated value (CSV) format. For the second process, data is analyzed using the proposed measurement tool, which includes both automated and semi-automated data analysis. Using Weka, a data mining tool, and MATLAB, the collected data are labeled with activity types, based on specific data patterns.

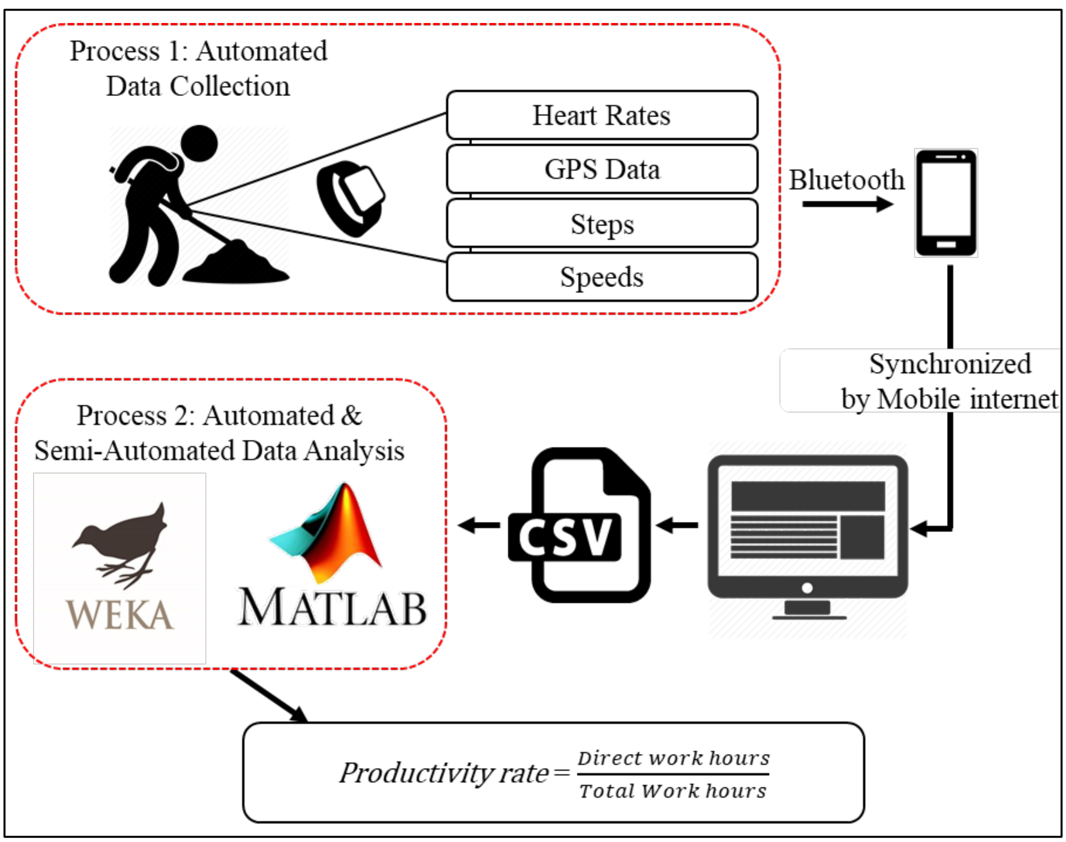


Figure 6. Semi-automated productivity measurement framework.

3.2. DATA COLLECTION

For this research, one of primary focuses was automatic data collection from laborers. The data collection was implemented by a wrist activity tracker, focusing on heart rates, steps, GPS data, and speed. Additionally, pilot testing projects were conducted to demonstrate the feasibility of the proposed data collection.

The author conducted three different types of data collection simultaneously at a construction site during field tests to compare the productivity rates of all three collection methods. The three types of data collection were as follows: 1. Data collection by an activity tracker, 2. Video data collection, 3. Manual data collection. The purpose of video collection and manual data collection is to get ground truth that can be compared with the data collected by a wrist activity tracker. Once the activity tracker was placed on the wrist of a laborer and data collection began, the other two data collection methods began 10 minutes later, since the activity tracker requires time to adjust its sensor settings, based on the individual user.

Automated data collection by a wrist activity tracker

An important first step is to provide a sufficient, reasonable explanation to the laborer so that they are willing to join in using the wrist activity tracker. The productivity measurements collected by the wrist activity tracker reports on an individual's productivity. Concerns by workers about individual productivity being measured could result in an unwelcome atmosphere at the construction site to use of a tracking system. To persuade laborers to accept the use of a tracking device, management should provide benefits to workers for compliance.

After obtaining agreement from a worker to be monitored, the activity tracker, Garmin Vivoactive HR, is put on the left wrist for a right-hander or the right wrist for a left-hander. The heart rate sensor is attached to the surface of the wrist on the other side of palm. To avoid measurement errors from the sensor, the tracker should not be tied loosely; the Garmin Optical Sensor check heart rates by sensing expansion and constriction of blood vessels inside the wrist just below the sensor. Then, its activity tracking application, which records heart rates, the standard time, speed, distance, steps, and GPS data, should be activated. At the end of the measurement, the application should be turned off. The start and finish of the application are done with a few touches of the screen and buttons.

Video Data Collection for Ground Truth

For video data collection, the author used a 12-megapixel camera of IPhone 7. The camera is shown fixed on only the author's chest by a chest mount for ground truth. The activities of a laborer wearing the activity tracker are recorded throughout the data collection of the activity tracker. The video recorded is used to validate the data collected by a wrist activity tracker.

Figure 7. Chest strap for fixed camera of the phone (An image adapted from Amazon).

Manual Data Collection for Ground Truth

In addition to the video data collection, the activities of a laborer are manually recorded on activity sheets shown in Figure 8 simultaneously with the automated data collection and video data collection (Figure 9.a). Blank areas used for handwritten recording of the labor's activity in the sheet are spaced at interval of 30 seconds.

Activity Observation Sheet

Wonseuk Son

12/24/2016

Age: Gender: Experience: Height: Weight: Temp:

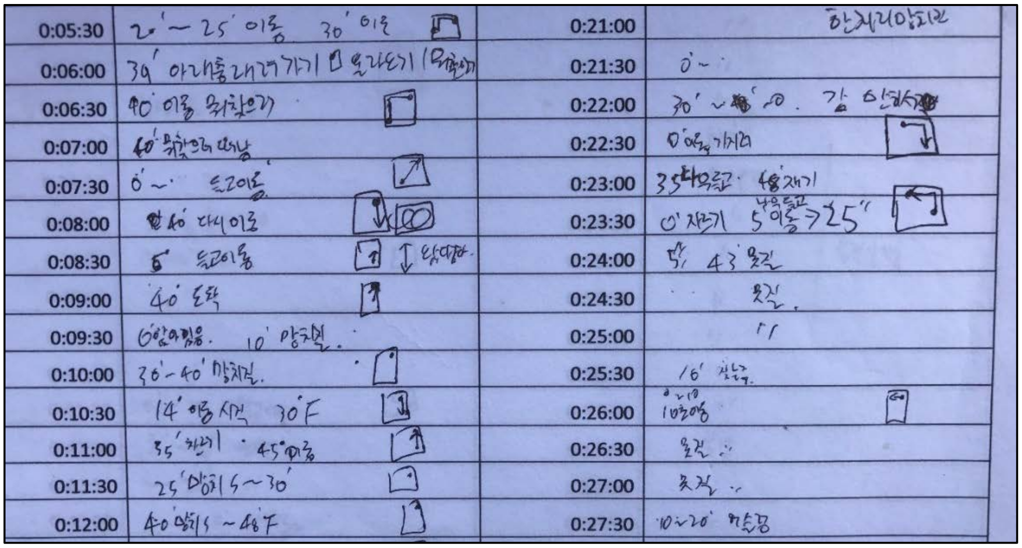
Start time:

TIME	ACTIVITY	0:15:00	
0:00:00		0:15:30	
0:00:30		0:16:00	
0:01:00		0:16:30	
0:01:30		0:17:00	
0:02:00		0:17:30	
0:02:30		0:18:00	
0:03:00		0:18:30	
0:03:30		0:19:00	
0:04:00		0:19:30	
0:04:30		0:20:00	
0:05:00		0:20:30	
0:05:30		0:21:00	
0:06:00		0:21:30	
0:06:30		0:22:00	
0:07:00		0:22:30	
0:07:30		0:23:00	
0:08:00		0:23:30	
0:08:30		0:24:00	
0:09:00		0:24:30	
0:09:30		0:25:00	
0:10:00		0:25:30	
0:10:30		0:26:00	
0:11:00		0:26:30	
0:11:30		0:27:00	
0:12:00		0:27:30	
0:12:30		0:28:00	
0:13:00		0:28:30	
0:13:30		0:29:00	
0:14:00		0:29:30	
0:14:30		0:30:00	

Figure 8. Activity observation sheet.

The author recorded the start and finish of the laborer's work activity in the blank form, measured in 30-second intervals. For example, when the worker wearing the wrist activity tracker started hammering at 1:30:10 and finishes at 1:30:30, the author recorded the start and finish in the 1:30:00 blank with each second such as, S(tart) hammering 10" and F(inish) 30" (Figure 9.b).

a. Manual activity recording at a construction site.



b. Ground truth recorded activity sheet (in Korean).

Figure 9. Sample of handwritten data recording by the author.

3.3. DATA PREPROCESSING

Raw data from the activity tracker is provided in web-based dataset that is exported in a comma-separated values format for interoperability in different software (Figure 10). It consists of 36 attributes, such as heart rates, GPS data, speed, etc. in columns and the instances (=rows) are recorded per second.

	ns1:Time	ns1:LatitudeDegrees	ns1:LongitudeDegrees	ns1:AltitudeMeters	ns1:Distance	ns1:Value4	Distance	ns2:Speed	(	ns2:Av	ns2:Av	ns2:	ns1:Name	ns1:UnitId
13	2016-12-27T00:33:20.	37.31498091	127.0731106	97.59999847	2.119999886	94	0.429999828	0.009	0	0.318	17	77	vivoactive HR	393215144
14	2016-12-27T00:33:21.	37.31498326	127.0731076	97.40000153	2.5	93	0.380000114	0.009	0	0.318	17	77	vivoactive HR	393215144
15	2016-12-27T00:33:22.	37.31498452	127.0731045	97.19999695	2.700000048	93	0.200000048	0.009	0	0.318	17	77	vivoactive HR	393215144
16	2016-12-27T00:33:23.	37.31498636	127.0731026	96.80000305	2.700000048	93	0	0.009	0	0.318	17	77	vivoactive HR	393215144
17	2016-12-27T00:33:24.	37.31498812	127.0731003	97.19999695	2.829999924	93	0.129999876	0.009	0	0.318	17	77	vivoactive HR	393215144
18	2016-12-27T00:33:25.	37.31498921	127.0731005	97.40000153	2.829999924	93	0	0.009	0	0.318	17	77	vivoactive HR	393215144
19	2016-12-27T00:33:26.	37.31498863	127.0731014	97.59999847	2.829999924	93	0	0.009	0	0.318	17	77	vivoactive HR	393215144
20	2016-12-27T00:33:27.	37.31499055	127.0731012	97.40000153	3.00999999	93	0.180000067	0.009	0	0.318	17	77	vivoactive HR	393215144
21	2016-12-27T00:33:28.	37.3149898	127.0731027	97.40000153	3.00999999	93	0	0.009	0	0.318	17	77	vivoactive HR	393215144
22	2016-12-27T00:33:29.	37.31498871	127.073104	97.40000153	3.00999999	93	0	0.009	0	0.318	17	77	vivoactive HR	393215144
23	2016-12-27T00:33:30.	37.31498854	127.0731058	97.40000153	3.00999999	94	0	0.009	0	0.318	17	77	vivoactive HR	393215144
24	2016-12-27T00:33:31.	37.31498661	127.0731063	97.40000153	3.00999999	94	0	0.009	0	0.318	17	77	vivoactive HR	393215144
25	2016-12-27T00:33:32.	37.31498544	127.0731072	97.40000153	3.00999999	95	0	0.009	0	0.318	17	77	vivoactive HR	393215144
26	2016-12-27T00:33:33.	37.31499558	127.0731139	97.40000153	4.010000229	96	1.000000238	0.009	0	0.318	17	77	vivoactive HR	393215144
27	2016-12-27T00:33:34.	37.31499692	127.0731157	97.40000153	4.130000114	96	0.119999886	0.009	0	0.318	17	77	vivoactive HR	393215144
28	2016-12-27T00:33:35.	37.31500145	127.0731177	97.40000153	4.550000191	97	0.420000076	0.009	0	0.318	17	77	vivoactive HR	393215144
29	2016-12-27T00:33:36.	37.31500723	127.0731234	97.40000153	5.170000076	97	0.619999886	0.019	0	0.318	17	77	vivoactive HR	393215144
30	2016-12-27T00:33:37.	37.31501511	127.0731306	97.40000153	5.900000095	97	0.730000019	0.019	0	0.318	17	77	vivoactive HR	393215144
31	2016-12-27T00:33:38.	37.3150183	127.0731355	97.40000153	6.079999924	98	0.179999828	0.019	0	0.318	17	77	vivoactive HR	393215144
32	2016-12-27T00:33:39.	37.31501956	127.0731386	97.40000153	6.320000172	98	0.240000248	0.019	0	0.318	17	77	vivoactive HR	393215144
33	2016-12-27T00:33:40.	37.31501939	127.0731395	97.40000153	6.320000172	98	0	0.019	0	0.318	17	77	vivoactive HR	393215144
34	2016-12-27T00:33:41.	37.31501855	127.0731405	97.59999847	6.320000172	98	0	0.028	0	0.318	17	77	vivoactive HR	393215144
35	2016-12-27T00:33:42.	37.31501947	127.0731416	97.59999847	6.320000172	98	0	0.028	0	0.318	17	77	vivoactive HR	393215144
36	2016-12-27T00:33:43.	37.31501914	127.0731419	97.59999847	6.320000172	98	0	0.028	0	0.318	17	77	vivoactive HR	393215144
37	2016-12-27T00:33:44.	37.31501964	127.073142	97.59999847	6.320000172	98	0	0.028	0	0.318	17	77	vivoactive HR	393215144
38	2016-12-27T00:33:45.	37.31501981	127.0731426	97.59999847	6.320000172	98	0	0.028	0	0.318	17	77	vivoactive HR	393215144
39	2016-12-27T00:33:46.	37.31502207	127.0731426	97.59999847	6.320000172	98	0	0.028	0	0.318	17	77	vivoactive HR	393215144
40	2016-12-27T00:33:47.	37.3150219	127.0731422	97.59999847	6.349999905	98	0.029999733	0.028	0	0.318	17	77	vivoactive HR	393215144
41	2016-12-27T00:33:48.	37.31501972	127.0731422	97.59999847	6.349999905	98	0	0.028	0	0.318	17	77	vivoactive HR	393215144
42	2016-12-27T00:33:49.	37.31501905	127.0731436	97.59999847	6.409999847	98	0.059999943	0.028	0	0.318	17	77	vivoactive HR	393215144
43	2016-12-27T00:33:50.	37.3150178	127.0731445	97.59999847	6.409999847	98	0	0.019	0	0.318	17	77	vivoactive HR	393215144

Figure 10. Raw dataset transferred from the activity tracker.

Data Reduction

Only 4 attributes from the raw data are needed for data analysis: heart rates, speed, distances, and steps. The other 32 attributes out of 36 attributes are removed in Excel. In addition, both the first and last 1200 instances (1200 seconds= 20 minutes) are removed to avoid outliers that can exist during the adjustment time for sensor settings of the activity tracker and that can be caused by the laborer’s awareness of being observed. In the second experiment, the author did not remove the first and last 1200 instances. Instead, the author randomly extracted several partial sections from the raw data.

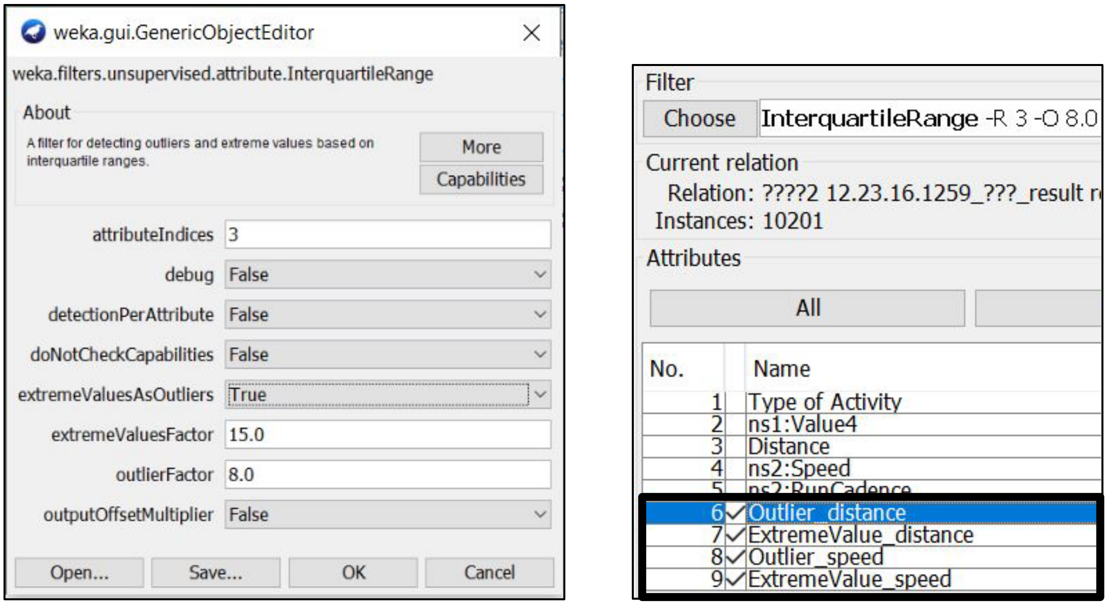
Data Cleaning

Even after data reduction, is it possible for the rough data to contain outliers. The outliers are usually created in the speed and distance sections that are detected and calculated by the accelerator and GPS, respectively. In terms of heart rates, it is difficult to find outliers in the rough data set. This is because the Garmin Optical HR sensor stops sensing when nothing is detected; the missing value is filled with the heart rate last measured.

The accelerator in a wrist activity tracker senses the speed of swings and movements of the wrist with the attached tracker. Some situations can generate extreme numerical values of speed. If a carpenter hammers nails with a hand wearing an activity tracker, the speed sensed by an accelerator will be much higher than when the carpenter is walking. Shaking hands with someone is also a factor that can create extreme values. Although it is recommended a wrist activity tracker not be worn on the user's dominant hand, there is no way to completely prevent outliers caused by the complexities of construction work in the hectic field of operations.

GPS data have another critical limitation. Enclosed work areas can block the GPS signal exchange between the device and a satellite, which leaves the GPS data with missing values. In addition, an obstacle placed over the device can distort the signal. The device can give extremely deviated GPS coordinates that make the distance moved per second data unreliable. Data cleaning is conducted only on the distance moved, not coordinates, because the input to be used is only for the distance moved. Coordinates are also referenced when deciding whether a laborer is walking to another location or moving around in circles by looking through track points. However, coordinates are not useful input for the performance measurement data tool.

Data cleaning is implemented with WEKA, a data mining tool. The first step is to detect outliers that have extreme values. Using the "InterquartileRange" filter in the WEKA (Figure 11.a), two attributes are added to the original dataset (Figure 11.b). Then, the added attributes are removed.



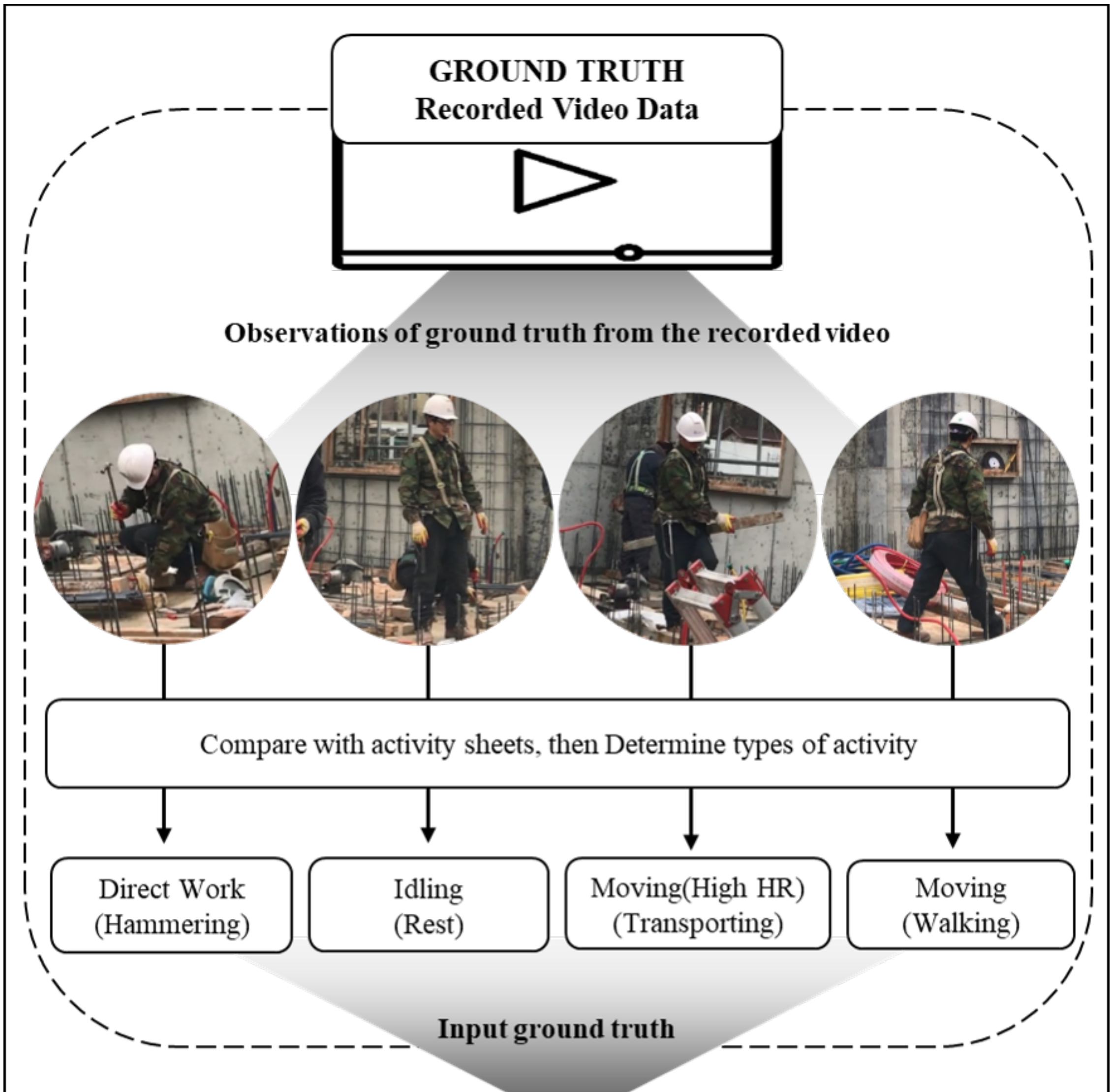
a. InterquatileRange Filter and Setting. b. Detected Outliers.

Figure 11. InterquatileRange filter and detected outliers.

Data Integration

In this research, data integration is the most crucial process for demonstration of the new automated data collection. Figure 12 illustrates the process of the data integration. Once all three datasets are collected from a laborer at a construction site, the types of activity observed from video data and the activity sheets (ground truth) must be input into the filtered data that is originally collected by a wrist activity tracker.

First, watching a recorded video, the author observes a laborer wearing a wrist activity tracker and determines the type of activity being performed. Then, the observations from the video are compared with the notes written on the activity sheets. Observations should be drawn in units of a second, according to the data of activity tracker that is also measured in seconds. After confirming the type of activity (ground truth), the activity types are entered into the activity tracker’s dataset.



Integrated into the data that is collected from a wrist activity tracker			
2016-12-23T04:22:27	Moving(High HR)	Transporting Timber	0.530
2016-12-23T04:22:28	Moving(High HR)	Transporting Timber	0.510
2016-12-23T04:22:29	Direct work	Hammering	0.000
2016-12-23T04:22:30	Direct work	Hammering	0.000
2016-12-23T04:22:31	Direct work	Hammering	0.000
2016-12-23T04:22:32	Idling	Standing & Rest	0.260
2016-12-23T04:22:33	Idling	Standing & Rest	0.170
2016-12-23T04:22:34	Moving	Walking	0.110
2016-12-23T04:22:35	Moving	Walking	0.040

Figure 12. Process of data integration of ground truth.

3.4. DATA ANALYSIS

Data analysis is performed using two methods. One method is automated analysis using classification algorithms of WEKA. Various classifications were implemented to obtain the most accurate results. The other method is semi-automated analysis. From a small portion of a dataset, the author manually detected specific patterns that respectively appeared in different activities observed from the video. The patterns were used for coding a MATLAB script to automatically assign the 4 types of activity to the whole dataset. Then, the newly assigned activities are compared with the actual activities observed from videos and activity data sheets.

3.4.1. Automated Data Analysis

Figure 13 illustrates the complete process of automated data analysis. Once the data integration finishes, the author implemented various classification algorithms in WEKA. Next, the author chose one of the most accurate classification methods to analyze the data. Finally, with results from the classification, the author compared the results with the result of the semi-automated data analysis and actual data observed from video and/or activity sheet, then provided interpretation.

The automated analysis was carried out using a decision tree classification method with a default setting (Figure 14) in WEKA. Based from several trials of basic data analysis at the experiments, the J48 decision tree classification provided the most accurate percentage of correctly classified instances.

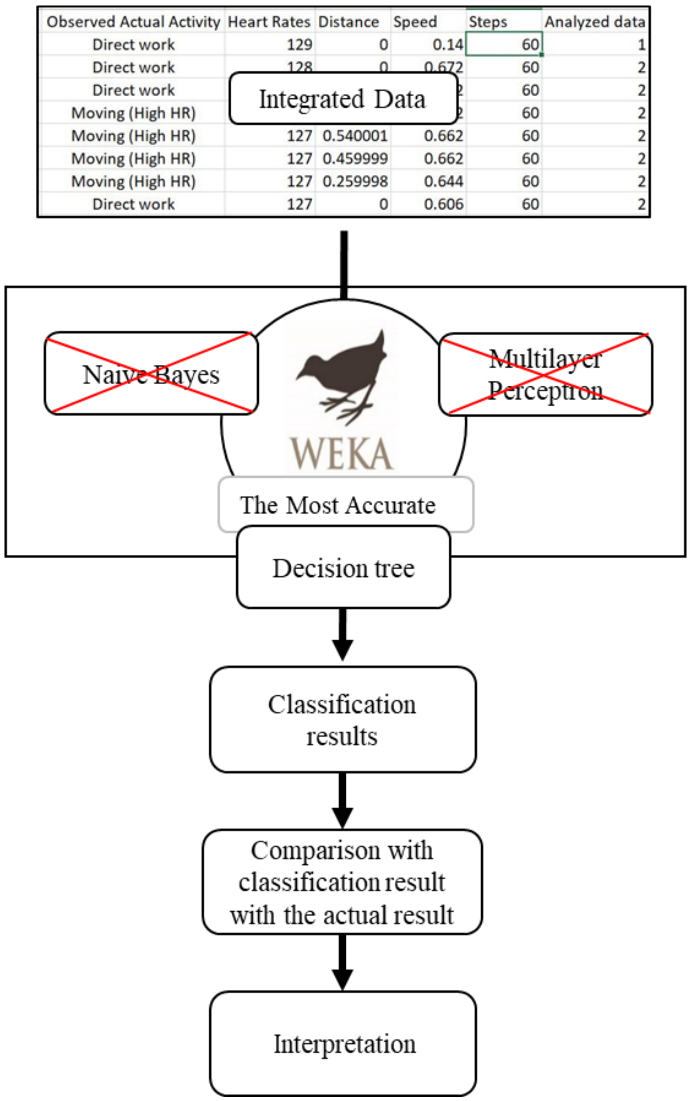


Figure 13. Process of automated data analysis.

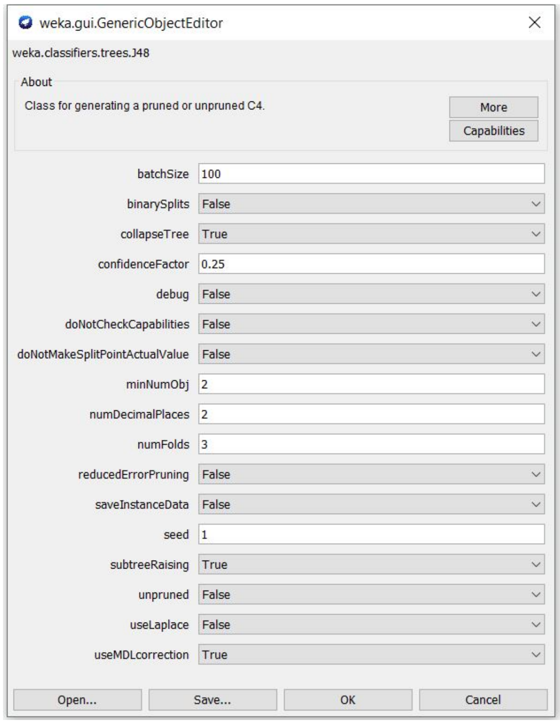


Figure 14. Default setting of J48 classification.

3.4.2. Semi-Automated Data Analysis

The semi-automated data analysis method consists of two major processes. One process is the manual recognition of patterns. The other process is the automated assignment of activity types by MATLAB. The purpose of the manual recognition of patterns is to catch some patterns that were missed in the automated data analysis. In the automated data analysis, relationships between values of the consecutive instances (=rows) recorded per second is not taken into account when WEKA detects/learns patterns from a training set. The values in a column are not independent data that has no correlation between the preceding and following values of an instance. The correlation between the consecutive values has a significant impact on detecting patterns, as some specific patterns can only be detected when looking at several subsequent values. Physical conditions and movements of laborers, such as speed, distance moved, and heart rates do not suddenly change in a second when the changes start or finish. The data usually have a gradual increase or decrease for several seconds. Therefore, it is significant to look at several consecutive values together in order to recognize an accurate start and finish point of a change in activity.

The process of semi-automated analysis begins with manual pattern detection. Once patterns are detected, the author codes the patterns into a MATLAB script, and then the script is operated to automatically assign a type of an activity to each instance (=row) of the integrated data according to observed patterns. Figure 15 illustrates the process.

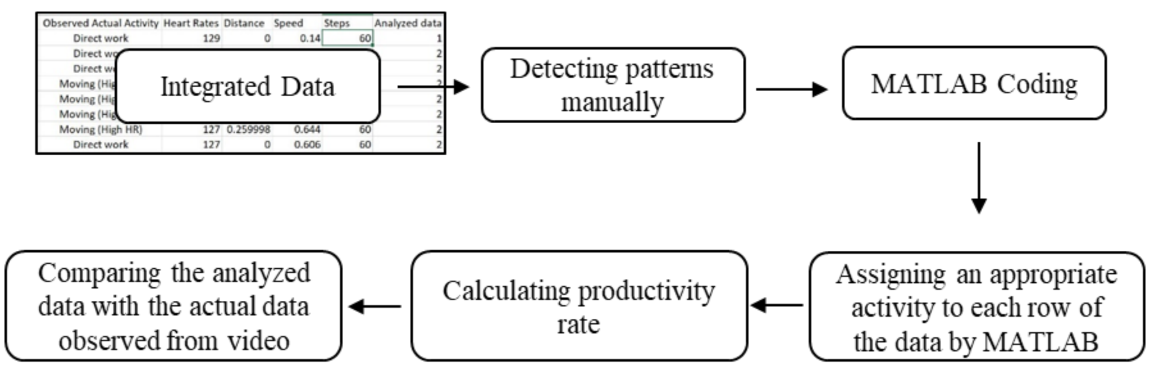


Figure 15. Process of semi-automated data analysis.

Manual Pattern Detection

The manual pattern detection is carried out by checking and understanding the values of heart rate, speed, moved distance, and steps shown in the 4 types of activity: Direct Work, Idling, Moving (High HR), and Moving. The expected basic patterns presented in the beginning of Chapter 3 are based on this manual detection. Only half of the data is used for the detection, which is a similar method of dividing a training set and a test set in WEKA.

There are two major criteria for pattern detection. One criterion is a specific value of heart rate that is used for division between resting heart rate and high heart rate. The other criterion is a combination of values of speed, steps, and moved distance for division between static movement and active movement. Using the two criteria, the four types of activity are classified in Figure 16.

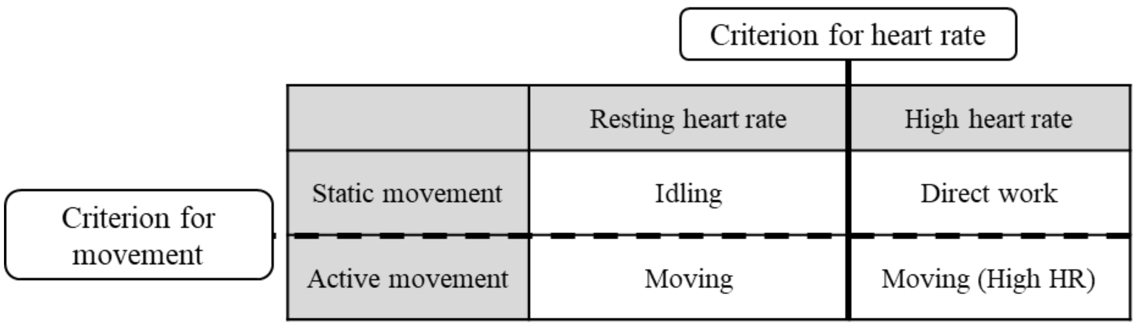


Figure 16. Activity assignment with manual pattern detection.

For direct work, the expected pattern is high heart rate with static/little movement. In contrast, for idling, a resting heart rate with static/little movement is expected. The concept of expected patterns for the two activities looks straightforward, because they have no movement. The challenge is to determine what specific heart rate is the criterion used to separate a resting heart rate from a high heart rate. Once the criterion heart rate is determined, it is applied to separate Moving and Moving (High HR) activity. Recognizing heart rate patterns between Moving and Moving (High HR), the author found that the criterion heart rate that is determined from the heart rate pattern of Idling and Direct work can be also used in sorting Moving and Moving (High HR).

Heart rate analysis

Heart rates requires a short period of time to increase or decrease to reach a stable heart rate, based on an average of regular heart rates that appear during a specific activity. Wrist activity trackers measure and record heart rate per second, so the heart rate cannot suddenly rise, for example, from 80 bpm (stable resting heart rate) to 110 bpm (stable high heart rate) in one second right after a laborer starts direct work (Figure 17. row 1). It should gradually increase from 80 bpm, 90 bpm, 100 bpm, to 110 bpm (Figure 17. row 2). This phenomenon is equally true in the opposite situation, when transitioning from direct work to idling status. It is statistically important to define a transition point of heart rate between 80 bpm and 110 bpm that will be used in MATLAB coding to sort resting heart rate and high heart rate.

Time	0:00:00	0:00:01	0:00:02	0:00:03	0:00:04	0:00:05
Heart rate (bpm)	80	80	~~80~~	~~110~~	110	110
Heart rate (bpm)	80	80	90	100	110	110

Figure 17. Transition of heart rates between resting heart rate and high heart rate.

Reviewing heart rate values during a transition range, the author determined a transition point of heart rate in the range. The simple method the author uses is to calculate and pick the average heart rate of the first and last heart rates in the range. In case of the above example, the transition point is 95 bpm, the average of 80 bpm and 110 bpm. However, there are numerous transition ranges in a sample of data provided by a wrist activity tracker. Therefore, the author was required to refer to as many transition ranges in a piece of data as possible until a solid transition point was reasonably determined (Figure 18).

Activity	Heart rate	Status of heart rate
Idling	80 ⋮ 80	Stable relaxing heart rate
Direct work	80	Transition range
	85	
	90	
	95	
	100	
	106	
	110	
	110 ⋮ 110	Stable high heart rate
Idling	112	Transition range
	105	
	100	
	96	
	90	
	86	
	80	
	79 ⋮ 79	Stable relaxing heart rate
Direct work	79	Transition range
	85	
	90	
	94	
	100	
	106	
	109	
	109 ⋮ 109	Stable high heart rate
Idling	110	Transition range
	105	
	100	
	95	
	90	
	86	
	80	
	80 ⋮ 80	Stable relaxing heart rate

A transition point determined from several transition ranges (95+96+94+95)=**95 bpm**

Figure 18. Example of determining a transition point of heart rate.

Movement data analysis

Regarding the movement data, the values must be interpreted by looking at GPS track points with numeric values because it is too difficult to determine movement patterns solely through numeric values in data. When a person begins to move, speed and moved distance gradually increase by acceleration. The person stops movement, the values gradually decrease. In particular, workers working at an active construction site require a few extra seconds to reach their normal moving speed because they are usually surrounded by equipment, materials, safety fences, or other workers. In these cases, the numeric values recorded at the start and finish can be regarded as static movement due to its low numeric values. In Figure 19, the track points are very close at the start and finish points while movement is in progress. The proximity between two track points show that the distance moved per second is very short and they have very low numeric values, similar to static movement.

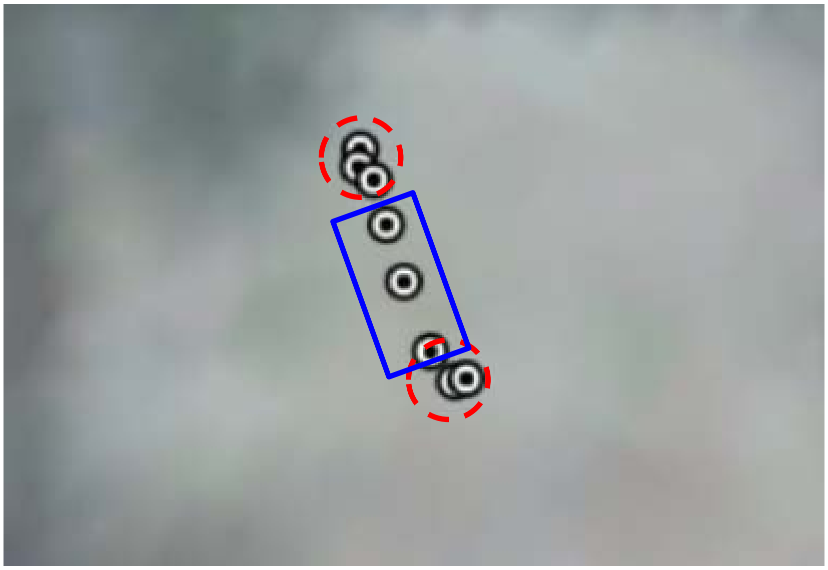


Figure 19. GPS Track points while walking.

Moreover, when a laborer walking in a forward direction changes direction to the opposite direction or moves at a sharp angle, the values of moving distance per second recorded while walking straight may become much shorter or zero at the turn. In Figure 20, the three red circles indicate turns and the blue boxes show movement walking forward. At the points of turn in the red circles, the two or three track points are close to each other. They can be also misclassified as static movements even if they are small parts of an active movement.

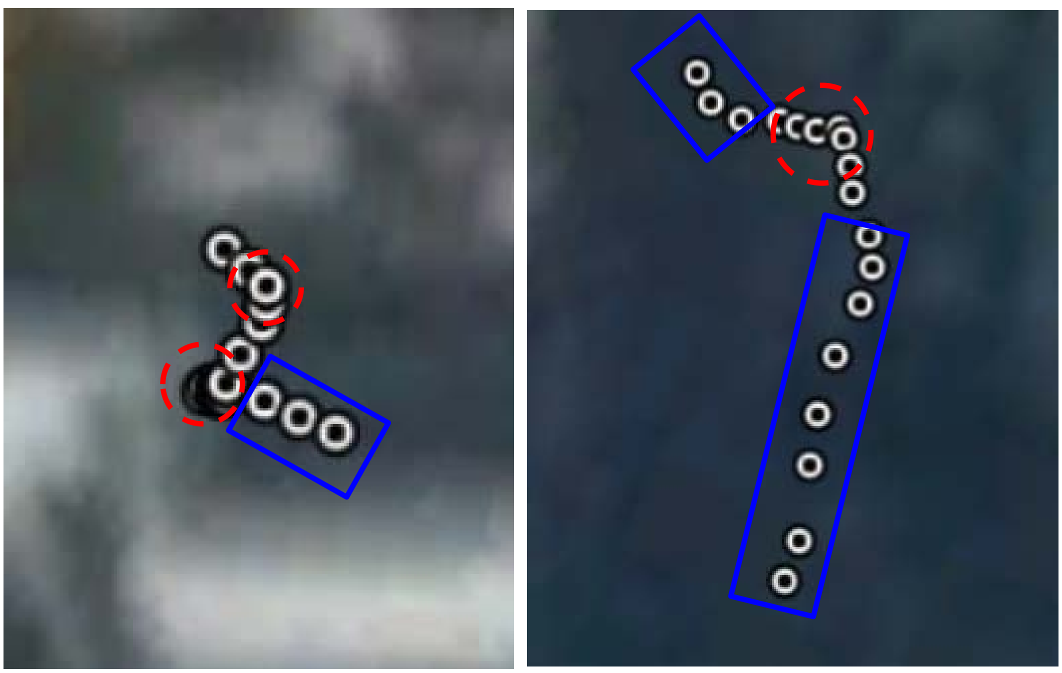


Figure 20. Close track points of turn while walking

Comparing actual static movements in Figure 21 with the above start/finish/turn track points in the red circles, the author believed that a method was needed to distinguish this data from static movement data. In MATLAB coding, after instances of active movement showing clear moving patterns are classified, the misleading track points can be differentiated by referring to previously classified activity types of their next and/or previous track points. For example, if the next and/or previous track points of a start/finish/turn track point are classified as active movement based on high values of movement data and the start/finish/turn track point hold a value at least over a specific low value, then the misleading track points should also be classified as active movement.

Figure 21. Track points of static movement.

Detected patterns from three experimental projects

From different datasets of three construction projects, the author conducted manual detection on half of the outlier-filtered data. The author mainly focused on finding the transition heart rate in which an instance is classified as a resting heart rate or a high heart rate. The patterns of movement values that were recorded at the start/stop and the sharp direction change of the carpenter were also carefully detected. Table 3 shows the patterns detected. From the 3 datasets collected from 3 carpenters at different construction sites, the author found that the only difference between the datasets was transition heart rate. Therefore all datasets were analyzed with the same patterns with different transition heart rate according to the laborers.

Transition heart rate	100 bpm (For example)
Type of Activity	**Patterns**
Moving (High HR)	1. Heart rate ≥ 100 bpm & Distance ≥ 0.3 & Distance of the next instance ≥ 0.26 2. Heart rate ≥ 100 bpm & Distance ≥ 0.3 & Distance of the previous instance ≥ 0.26 3. Heart rate ≥ 100 bpm & Distance ≥ 0.2 & the assigned activity of the previous instance = "Moving (High HR)" 4. Heart rate ≥ 100 bpm & Speed ≥ 0.5 & Step ≥ 40 5. Heart rate ≥ 100 bpm & Distance ≥ 0.3 & Speed ≥ 0.16
Direct work	The rest of instances meeting "Heart rate ≥ 100 bpm" after assignment of "Moving (High HR)"
Moving	Same patterns of movement as the "Moving (High HR)", under the condition "Heart rate < 100 bpm"
Idling	The rest instances meeting "Heart rate < 100 bpm" after assignment of "Moving"

Table 3. Manually detected patterns.

Automated Activity type assignment by MATLAB

The above detected patterns were coded on a MATLAB script by the author. Then, each dataset was analyzed by the script. Figure 22 shows the basic concept of automated assignment of the activity type. Figure 23 is the MATLAB script that the author implemented for the assignment. MATLAB is not efficient with string data values, so the activity types were represented and assigned as the number 0, 1, 2, and 3: “Idling”, “Direct Work”, “Moving (High HR)”, and “Moving” respectively.

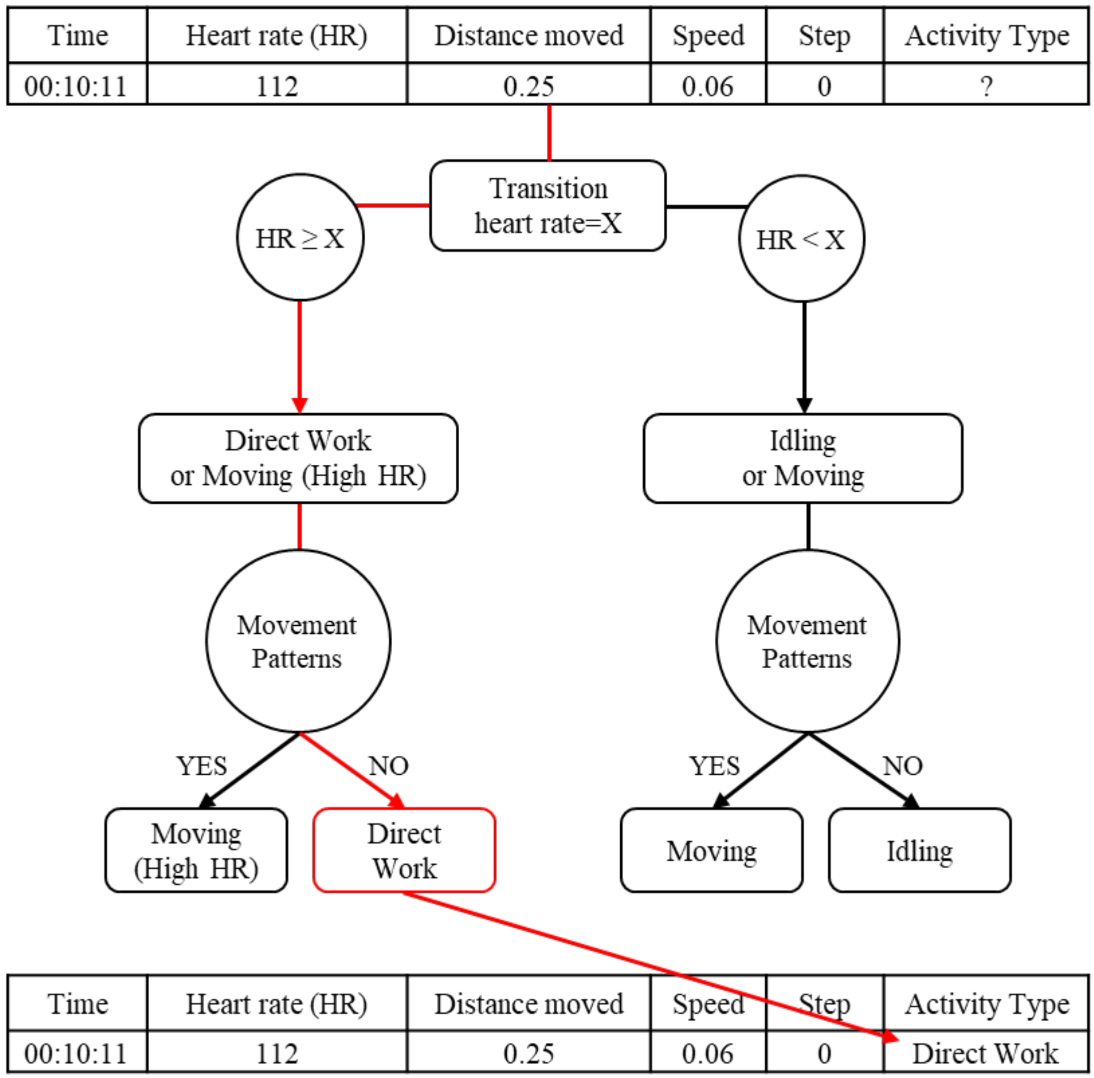


Figure 22. Example of activity type assignment by MATLAB.

```
clear;
clc;
close all

HR=raw_data(:,14); % Extract HR
SP=raw_data(:,15); % Extract SP
ST=raw_data(:,16); % Extract STEP

D1=raw_data(2:length(raw_data),13); %Distance 1 extracted
D2=raw_data(1:length(raw_data)-1,13); %Distance 2
D3=D1-D2; %Distance

F=[0]; %First row added
Ds=[F;D3]; %Distance
F_data=cat(2,HR,Ds,SP,ST);

for n=1:length(F_data);

    if HR(n,:)>=100 && Ds(n,:)>=0.3 && Ds(n+1,:)>=0.26;
            F_data(n,5)=2;
    elseif HR(n,:)>=100 && Ds(n,:)>=0.3 && Ds(n-1,:)>=0.26;
            F_data(n,5)=2;
    elseif HR(n,:)>=100 && Ds(n,:)>=0.2 && F_data(n-1,5)==2;
            F_data(n,5)=2;
    elseif HR(n,:)>=100 && SP(n,:)>=0.5 && ST(n,:)>=40;
            F_data(n,5)=2;
    elseif HR(n,:)>=100 && Ds(n,:)>=0.3 && SP(n,:)>=0.16;
            F_data(n,5)=2;
    elseif HR(n,:)>=100 && Ds(n,:)<0.2; % Heart rate>=100
            F_data(n,5)=1;
    elseif HR(n,:)>=100 ;
            F_data(n,5)=1;

    %%%%%%%%%%%%%%%%%%%%%%%Resting heart rate%%%%%%%%%%%%%%%%%%%%%%%%
    elseif HR(n,:)<100 && Ds(n,:)<0.2; % Heart rate>=100
            F_data(n,5)=0;

    elseif HR(n,:)<100 && Ds(n,:)>=0.3 && Ds(n+1,:)>=0.26;
            F_data(n,5)=3;
    elseif HR(n,:)<100 && Ds(n,:)>=0.3 && Ds(n-1,:)>=0.26;
            F_data(n,5)=3;
    elseif HR(n,:)<100 && Ds(n,:)>=0.2 && F_data(n-1,5)==2;
            F_data(n,5)=3;
    elseif HR(n,:)<100 && SP(n,:)>=0.5 && ST(n,:)>=40;
            F_data(n,5)=3;
    elseif HR(n,:)<100 && Ds(n,:)>=0.3 && SP(n,:)>=0.16;
            F_data(n,5)=3;
    else F_data(n,5)=0;
    end
end
```

Figure 23. MATLAB Script for activity type assignment.

Chapter 4: Field Demonstration of the Proposed Methodology

The proposed personal productivity measurement methodology was tested at three different construction sites in Seoul, South Korea. Two of the sites were residential buildings. The third site was a church construction project. The type of work selected was labor intensive so that heart rate data could show distinct patterns according to activities. In addition, the experiment had to be conducted outside due to the limitations of GPS, which transmits too low a signal in an enclosed environment. Therefore, the data collection portion of the research was implemented on labor-intensive formworks outside. Three carpenters joined the experiment.

4.1. RESIDENTIAL BUILDING PROJECT 1

The first experiment was conducted at a construction project of a new residential building. The major work was concrete formwork on the top floor. The laborer was a carpenter with 25 years of experience. He worked primarily on the formwork for the elevator wall shown in Figure 24. Detailed information on the project and experiment is described in Table 4.

Figure 24. Formwork for elevator wall by a carpenter.

Contents	Description
Type of project	● New residential building construction
Location	● Seoul, South Korea
Major work	● Concrete Formwork for elevator wall
Observation time (Video)	● 3 hours, 30 Minutes
Number of data collected (Seconds)	● 12,601 (Raw) 10,174 (Analyzed)
Transition heart rate	● 100 bpm
The detail of activity types	● Direct work: Hammering, Assembling forms, and Cutting/Sawing woods.
	● Idling: Calling, Chatting, Waiting Standing & Sitting with no motion. Smoking, and Drinking water.
	● Moving: Walking, Loitering, And Traveling.
	● Moving (High HR): Transporting material, a little movement during direct works.

Table 4. Detailed information of the project and experiment.

4.1.1. Data Collection & Preprocessing

Video was recorded for 3 hours, 30 minutes, and 12,601 pieces of data, including heart rate, speed, steps, and distance moved, were collected by a wrist activity tracker from a carpenter working on formwork. In addition, the track points from the GPS were also successfully marked and are shown in Figure 25.

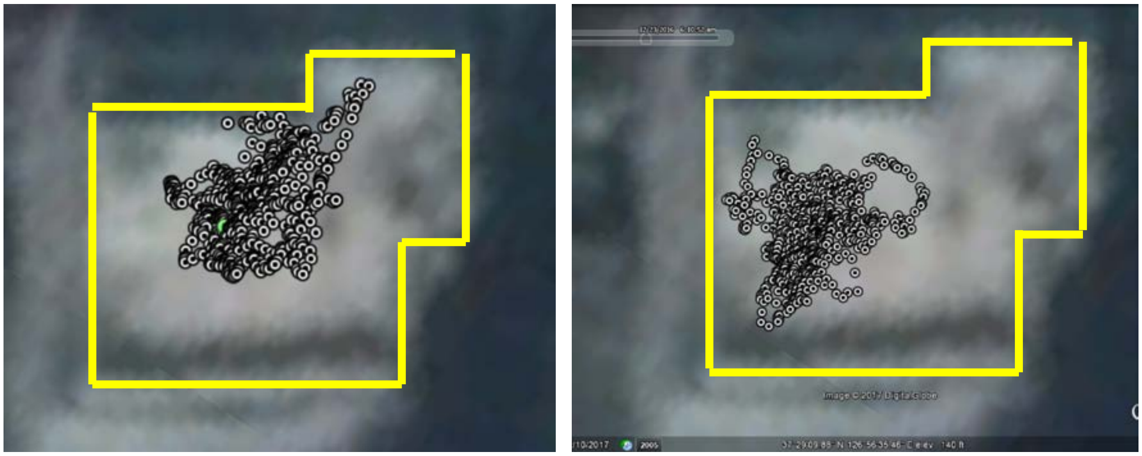


Figure 25. Track points collected from GPS.

A total of 1200 pieces (20 minutes) of data were removed from each of the first instance (row) and the last instance (row): a total of 2400 pieces of data (40 minutes). WEKA analysis detected and removed 27 outliers, so the total number of data instances used was 10,174.

4.1.2. Results of Automated Data Analysis

The percentage of correctly classified instances by Decision tree algorithm was 85.79%. Table 5.a shows the detail of the classification accuracy. The most important parts, “Direct Work” and “Idling”, had a good result in classification as a true positive rate, precision, and recall, but 27.9% of other instances were misclassified as “Direct Work” (FP rate). The decision tree algorithm correctly classified only 37.3% of actual “Moving (High HR)” instances (Recall). The “Moving (High HR)” had the lowest precision and recall rates. In Table 5.b, more than 50% of “Moving (High HR)” instances were misclassified as direct work, yet they were rarely classified as other types of activity. The reason is that movements at the site were very irregular due to the small area of the site, so it was difficult to observe linear movements. For example, when the carpenter moved 30 feet in direction, he had to stop and go, and several small turns were needed due to the congested, small construction area.

Class	TP Rate	FP Rate	Precision	Recall
Direct work	93.2%	27.9%	87.4%	93.2%
Moving (High HR)	37.3%	2%	70.4%	37.3%
Idling	87.6%	3.9%	82.7%	87.6%
Moving	89.5%	0.2%	95.4%	89.5%

a. Detailed Accuracy by J48 Classification of the residential building project 1.

		Predicted activity			
		Direct work	Moving (High HR)	Idling	Moving
Actual Activity	Direct work	6395	173	293	3
	Moving (High HR)	705	425	6	4
	Idling	210	4	1577	9
	Moving	7	2	30	331
Total		7317 (72%)	604 (6%)	1906 (19%)	347 (3%)

b. Confusion Matrix by J48 Classification of the residential building project 1.

Table 5. Results of J48 classification on the residential building project 1.

4.1.3. Results of Semi-Automated Data Analysis

The MATLAB script was coded by the author according to the patterns successfully performed on the collected data. An activity type was assigned to each instance by number:

0-Idling 1-Direct work, 2-Moving (High HR), and 3-Moving. The last column in Table 6 shows the activity types assigned by MATLAB. The transition heart rate was determined to be 100 bpm. The result of the semi-automated assignment of activity types is shown in Table 7.

1:32:06	Idling	864.7199707	97	0.090	0.14	60	0
1:32:07	Idling	864.8499756	98	0.130	0.14	0	0
1:32:08	Idling	864.8499756	99	0.000	0.14	0	0
1:32:09	Idling	864.9699707	99	0.120	0.14	0	0
1:32:10	Idling	865.039978	100	0.070	0.14	0	0
1:32:11	Idling	865.039978	100	0.000	0.14	0	1
1:32:12	Direct work	865.1699829	102	0.130	0.14	0	1
1:32:13	Direct work	865.1699829	102	0.000	0.14	60	1
1:32:14	Direct work	865.1699829	104	0.000	0.14	60	1
1:32:15	Direct work	865.1699829	106	0.000	0.14	60	1
1:32:16	Direct work	865.2600098	107	0.090	0.14	60	1
1:32:17	Direct work	865.3099976	107	0.050	0.149	60	1
1:32:18	Direct work	865.3200073	107	0.010	0.149	0	1
1:32:19	Direct work	865.3699951	107	0.050	0.149	0	1
1:32:20	Moving (High HR)	865.9299927	106	0.560	0.149	0	2
1:32:21	Moving (High HR)	866.25	106	0.320	0.149	0	2
1:32:22	Moving (High HR)	867.0499878	105	0.800	0.149	0	2
1:32:23	Moving (High HR)	867.4899902	104	0.440	0.149	0	2
1:32:24	Moving (High HR)	867.8400269	108	0.350	0.149	0	2
1:32:25	Moving (High HR)	868.0800171	106	0.240	0.149	0	2
1:32:26	Moving (High HR)	868.7600098	105	0.680	0.149	0	2

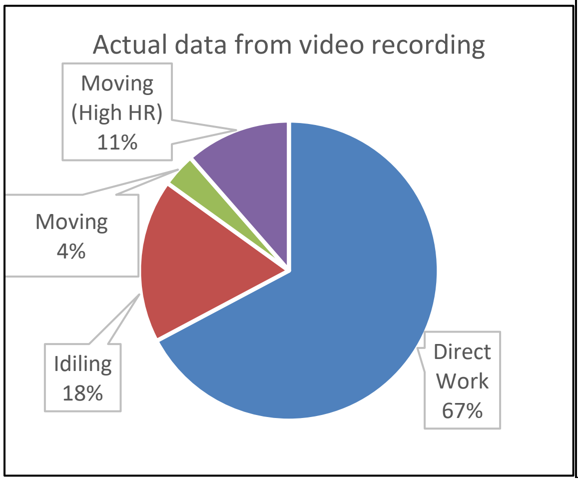
Table 6. A part of the result from activity assignment by MATLAB.

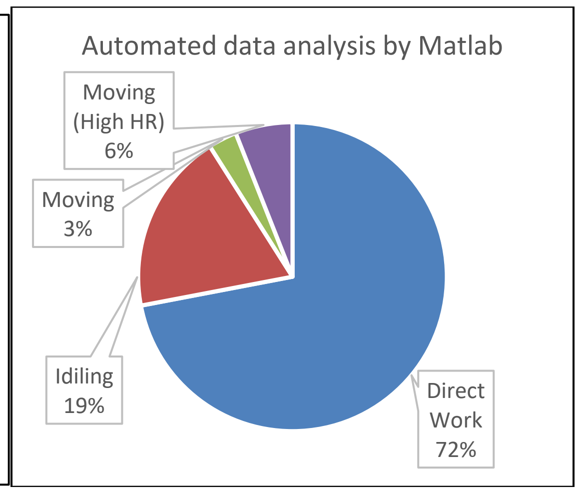
Activity type	Number of assigned instances
Direct Work	6207 (61%)
Moving (High HR)	1468 (14%)
Idling	2129 (21%)
Moving	370 (4%)
Total	10174

Table 7. Result of semi-automated data analysis of the building project 1.

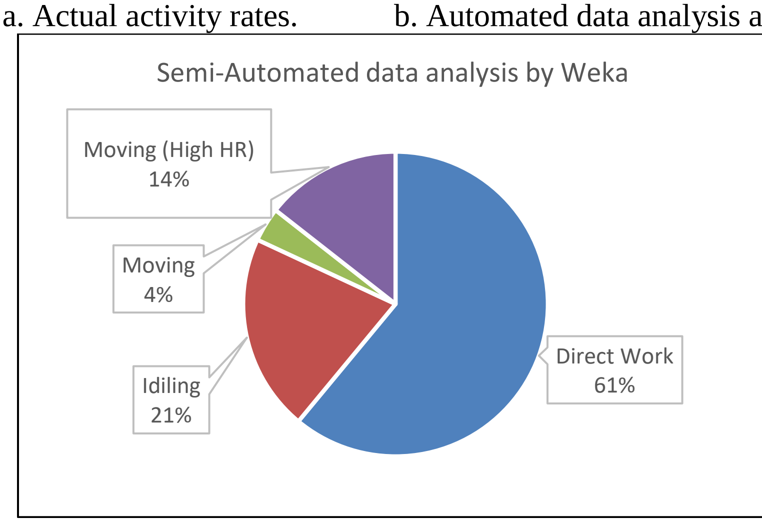
4.1.4. Work Rates Comparisons

Figure 26 illustrates the three different activity rates of the actual, automated, and semi-automated data analysis. All of the three pie charts have very similar patterns. The percentage of "Direct work" accounts for around 60-70% in all charts. Comparing the actual data with the two proposed data analysis methods, the results were fairly accurate, only having -5% to 6% difference in the all activity rates, as shown in Table 8.





a. Actual activity rates. b. Automated data analysis activity rates.



c. Activity rates of the semi-automated data analysis.

Figure 26. Results of the three methods in the Building Project 1.

	Actual data (a)	Automated method (b)	Semi-Automated Method (c)	Difference (a)-(b)	Difference (a)-(c)
Direct Work	6841 (67%)	7317 (72%)	6207 (61%)	-5%	6%
Moving (High HR)	1163 (11%)	604 (6%)	370 (14%)	5%	-3%
Idling	1800 (18%)	1906 (19%)	2129 (21%)	-1%	-3%
Moving	370 (4%)	347 (3%)	1468 (4%)	1%	0%
Total time (Seconds)	**10174 (100%)**	**10174 (100%)**	**10174 (100%)**		

Table 8. Comparisons between the three methods in the building project 1.

4.2. CHURCH PROJECT

The second experiment was conducted at a new construction project of a church. A carpenter with 20 years of experience joined the experiment. The main task was formwork of the first-floor exterior wall, as shown in Figure 27. The carpenter worked outdoor while collecting data. The detailed information is provided in Table 9.

Figure 27. Formwork for exterior wall.

Contents	Description
Type of Project	● New church construction
Location	● Seoul, South Korea
Major work	● Concrete formwork for exterior wall
Observation time (Video)	● 2 hours, 15 Minutes
Number of data collected (Seconds)	● 8,119 (Raw) ⟶ 3,973 (Analyzed)
Transition heart rate	● 110 bpm
The detail of activity types	● Direct work: Hammering, Assembling forms, and Removing the forms from finished concrete wall.
	● Idling: Calling, Chatting, Waiting, Standing & Sitting with no motion, and Drinking water.
	● Moving: Walking, Loitering, and Traveling.
	● Moving (High HR): Transporting material, a little movement during direct works.

Table 9. Detailed information of the church project and experiment.

4.2.1. Data Collection & Preprocessing

A total of 8,119 pieces of data were collected during 2 hours and 15 minutes of observation. Figure 28 shows clear track points. The construction site was a long rectangular shape, so the patterns of movement are more clearly shown in Figure 9, compared to the previous experimental project.

Figure 28. Track points collected from GPS in the church project.

In this experiment, the author randomly extracted three sections from the raw data. In total, 3,973 instances were extracted from three sections, accounting for half of the data. The three sections were extracted from the middle of data collected, so the first and last 2400 instances that could contain errors due to the adjustment processing of sensors in the device were not taken into account. Outliers were not detected from the extracted data by WEKA analysis.

4.2.2. Results of Automated Data Analysis

Overall, the J48 classification provided a good performance for three classes: “Direct work”, “Idling”, and “Moving”. In the three classes, a greater than 80% accuracy was achieved in the true positive, precision, and recall sections (Table 10.a). In addition, the most interesting observation for the church construction site dataset is that the “Moving (High HR)” class showed a higher result of accuracy, around 70%, in the true positive, precision, and recall section. When compared with the other two experiments, the 70% accuracy is significantly high. The reason for the comparatively high accuracy is that the shape of construction site was a long rectangle and the path that the carpenter moved and transported material through was clear enough for uninterrupted forward movement. The straight movement can be easily recognized with the track points in Figure 29. Table 10.b shows the result of J48 classification in this experiment. Although the accuracy of the “Moving (High HR)” was high compared to other experiments, 375 out of 1174 instances were misclassified as “Direct work”.

Class	TP Rate	FP Rate	Precision	Recall
Direct work	88.1%	22.3%	81.5%	88.1%
Moving (High HR)	67.1%	8.1%	77.7%	67.1%
Idling	91.8%	1.4%	92.1%	91.8%
Moving	81%	0.4%	86.4%	81%

a. Detailed accuracy by J48 classification of the church project.

		Predicted activity			
		Direct work	Moving (High HR)	Idling	Moving
Actual Activity	Direct work	1842	213	33	3
	Moving (High HR)	375	788	3	8
	Idling	42	1	534	5
	Moving	2	12	10	102
Total		2261 (57%)	1014 (25%)	580 (15%)	118 (3%)

b. Confusion matrix by J48 classification of the church project.

Table 10. Results of J48 classification on the church project.

4.2.3. Results of Semi-Automated Data Analysis

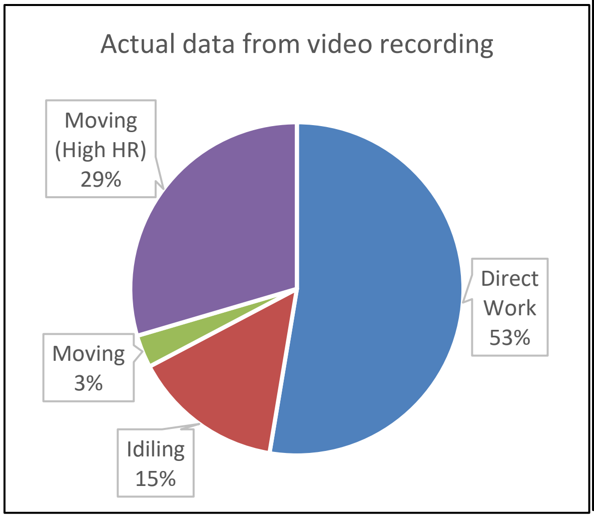
Table 11 shows the result of the semi-automated assignment of activity types by MATLAB. The overall percentage of the number of classified instances is highly comparable with the automated data analysis. The largest difference is detected in the "Direct work" and "Moving (High HR)" categories. The misleading movement data that was explained in "3.4.2. Manual Pattern Detection" could be the reason for that difference.

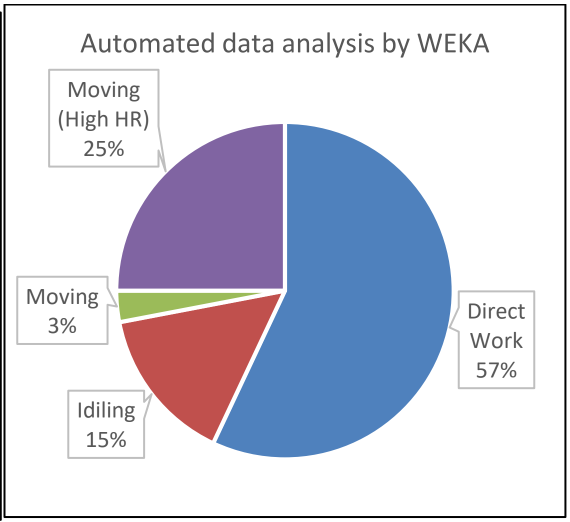
Activity type	Number of assigned instances
Direct Work	2021 (51%)
Moving (High HR)	1222 (31%)
Idling	584 (15%)
Moving	146 (4%)
Total	3973

Table 11. Result of semi-automated data analysis of the church project.

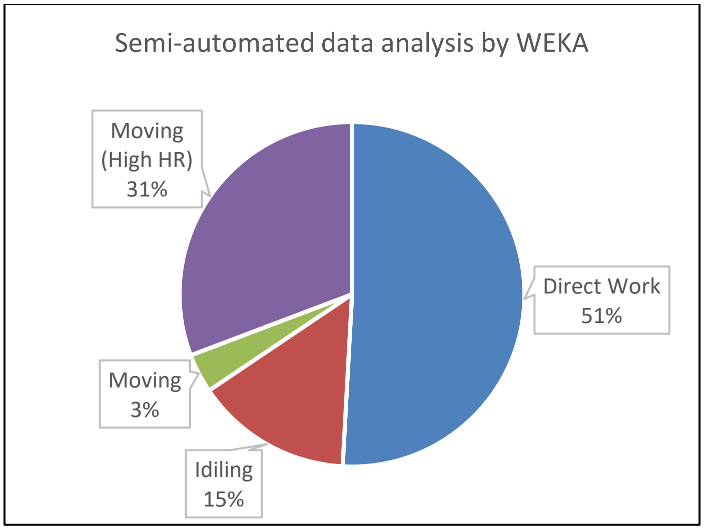
4.2.4. Work Rates Comparisons

The three pie charts in Figure 29 look very similar. In Table 12, the semi-automated data analysis shows a smaller difference, ranging from -1% to 2%. The largest difference is in the “Moving (High HR)” of the automated method, 5%. Overall, the difference is small: under ± 5% in all categories.





a. Actual activity rates. b. Automated data analysis’s activity rates.



c. Activity rates of the semi-automated data analysis.

Figure 29. Results of the three methods in the church project.

	Actual data (a)	Automated method (b)	Semi-Automated Method (c)	Difference (a)-(b)	Difference (a)-(c)
Direct Work	2091 (53%)	2261 (57%)	2021 (51%)	-4%	2%
Moving (High HR)	1174 (30%)	1014 (25%)	1222 (31%)	5%	-1%
Idling	582 (14%)	580 (15%)	584 (15%)	-1%	-1%
Moving	126 (3%)	118 (3%)	146 (4%)	0%	-1%
Total time (Seconds)	**3937 (100%)**	**3937 (100%)**	**3937 (100%)**		

Table 12. Comparisons between the three methods in the church project.

4.3. Residential Building Project 2

The last field test of the new productivity measurement was for the new construction of a residential building. The task was concrete formwork for an underground parking ramp, as shown in the Figure 30. A carpenter who has more than 30 years of experience joined this experiment. The formwork was for a cylindrical shaped ramp. Due to the curved surface, it was much more difficult for the carpenter to assemble and adjust forms. Table 13 is detailed information of this experiment.

Figure 30. Formwork of underground parking ramp.

Contents	Description
Type of Project	● New residential Building
Location	● Seoul, South Korea
Major work	● Concrete formwork for underground parking ramp
Observation time (Activity sheet)	● 2 hours, 20 Minutes
Number of data collected (Seconds)	● 8,402 (Raw) ⟶ 6,000 (Analyzed)
The detail of activity types	● Direct work: Hammering, Assembling forms, Sawing/ Cutting wood forms

	● Idling: Calling, Chatting, Waiting, Standing & Sitting with no motion, Eating a snack, Drinking water.
	● Moving: Walking, Loitering, and Traveling.
	● Moving (High HR): Transporting material, a little movement during direct works.

Table 13. Detailed information of the residential building project 2 and experiment.

4.3.1. Data Collection & Preprocessing

In this experiment, video recording was not allowed at the construction site, so manual observations were recorded on the activity sheet in as much detail as possible. Even when laborers moved to another area, the locations were recorded on the sheets. Location data was successfully collected from GPS with the track points shown in Figure 31. Transporting material was primarily accomplished by means of a tower crane, and the material was laid down and made ready on the work area just behind the carpenters. Most of the outliers that were detected in the movement data were not identified, ranging between normal values. The first and last 1200 instances, a total of 2400 instances, were removed in order to reduce adjustment errors in the sensors in the wrist activity tracker.

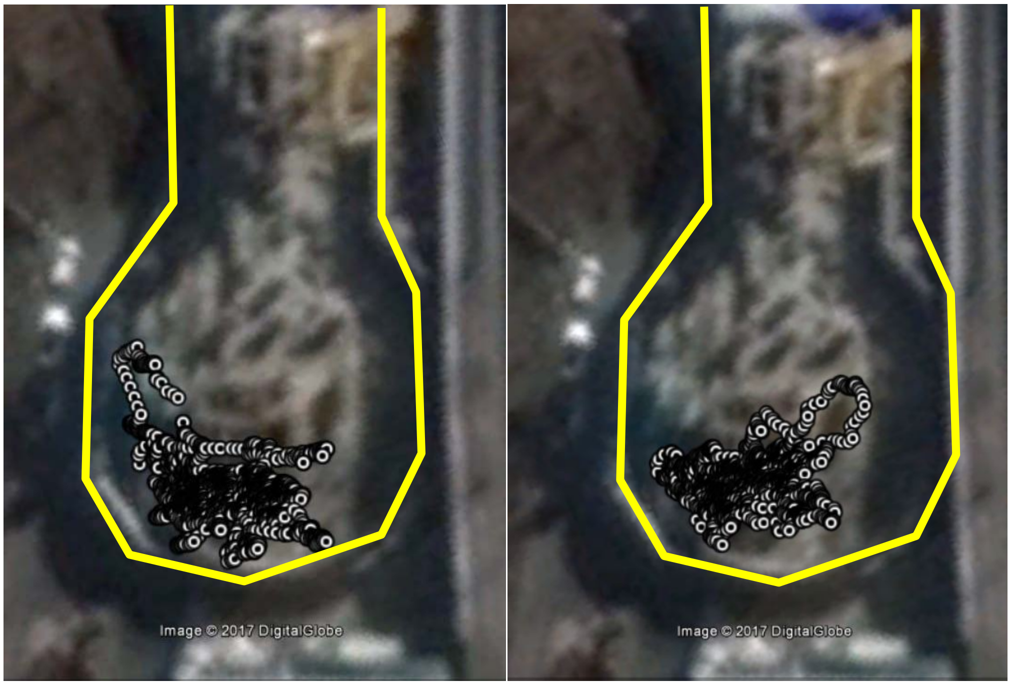


Figure 31. Track points recorded in the building project 2.

4.3.2. Results of Automated Data Analysis

The automated data analysis performed very well in the section "Direct work" and "Idling" with greater than 85% of true positive rate, precision and recall, as shown in Table 14.a. However, it showed a poor result in the two movement activities, providing less than 50% of true positive rate and less than 60% of precision. Approximately 5500 of 6000 instances (90%) were classified as "Direct work" or "Idling". But 172 out of 367 "Moving (High HR) instances were classified as "Direct work", shown in the Table 14.b, which was the poorest automated analysis classification data outcome in this experiment.

Class	TP Rate	FP Rate	Precision	Recall
Direct work	89%	9.7%	90.5%	89%
Moving (High HR)	44.1%	2.2%	56.8%	44.1%
Idling	92.3%	9.9%	85.1%	92.3%
Moving	47.9%	1.7%	59.6%	47.9%

a. Detailed accuracy by J48 classification of the residential building project 2.

		Predicted activity			
		Direct work	Moving (High HR)	Idling	Moving
Actual Activity	Direct work	2730	110	216	12
	Moving (High HR)	172	162	19	14
	Idling	102	5	2097	69
	Moving	11	8	133	140
Total		3015 (50%)	285 (5%)	2465 (41%)	235 (4%)

b. Confusion matrix by J48 classification of the residential building project 2.

Table 14. Results of J48 classification on the building project 2.

4.3.3. Results of Semi-Automated Data Analysis

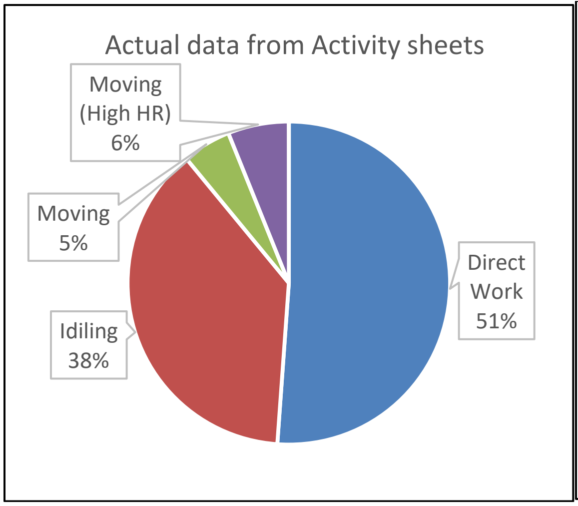
The transition heart rate was determined to be 101 bpm. The movement patterns detected and coded in the MATLAB were applied to this dataset. Table 15 is the result of the Semi-Automated Data analysis. The semi-automated data analysis shows almost the same percentage as found in the automated data analysis. Approximate 90% of the instances belong to the “Direct work” and “Idling”. This is because the laborer’s tasks were highly static work with little movement. All tools, equipment, and materials were ready at the work area. The interesting finding here is the “Idling” accounts for 41%, which is higher in comparison to the other two experiments. The reason for this was due to the type of work. The formwork was to assemble flat rectangular forms adjacent to a curved wall. The tasks were considered to be intensive work requiring a high level of concentration, and therefore the carpenter is allowed more break time.

Activity type	Number of assigned instances
Direct Work	2987 (50%)
Moving (High HR)	274 (4%)
Idling	2461 (41%)
Moving	278 (5%)
Total	6000

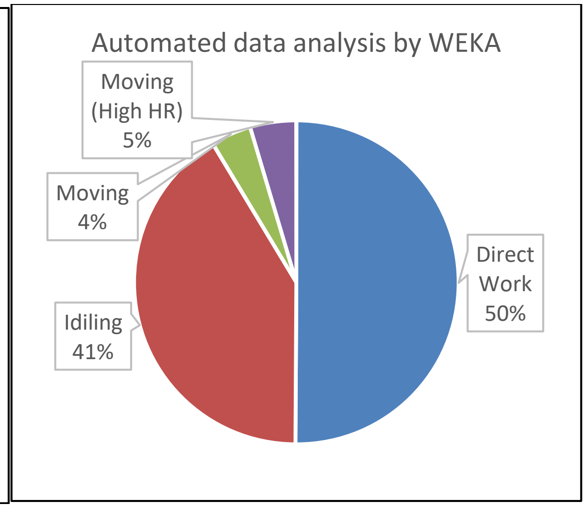
Table 15. Result of the semi-automated data analysis of the building project 2.

4.3.4. Work Rates Comparisons

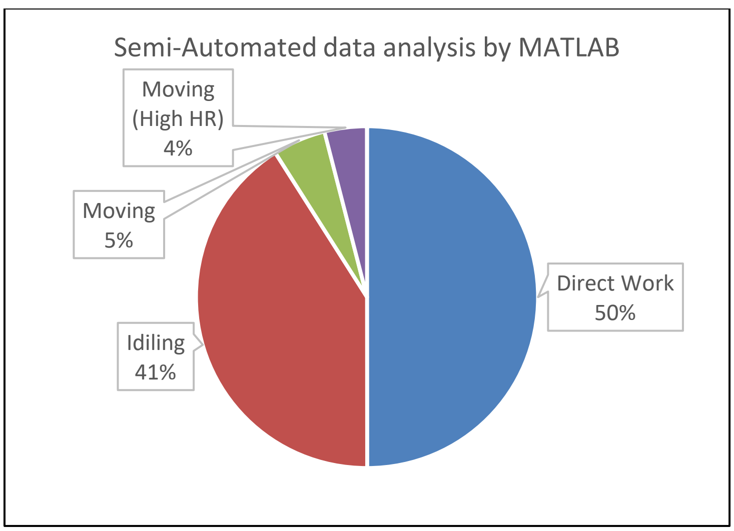
The result of the comparison of activity rates between the three datasets is shown in Figure 32 and Table 16. The differences are significantly low, ranging only ± 3%. The dominant percentages of “Direct work” and “Idling” are clearly shown in the three methods.



a. Actual activity rates.



b. Automated data analysis's activity rates.



c. Activity rates of the semi-automated data analysis.

Figure 32. Results of the three methods in the building project 2.

	Actual data (a)	Automated method (b)	Semi-Automated Method (c)	Difference (a)-(b)	Difference (a)-(c)
Direct Work	3068 (51%)	3015 (50%)	2987 (50%)	1%	1%
Moving (High HR)	367 (6%)	285 (5%)	274 (4%)	1%	2%
Idling	2273 (38%)	2465 (41%)	2461 (41%)	-3%	-3%
Moving	292 (5%)	235 (4%)	278 (5%)	1%	0%
Total time (Seconds)	**6000 (100%)**	**6000 (100%)**	**6000 (100%)**		

Table 16. Comparisons between the three methods in the building project 2.

Chapter 5: Conclusions

The main purpose of this research was to explore the feasibility of a personal productivity measurement that is based on labor productivity rate, which is defined as the ratio of direct work hours to total work hours. The scope focused on automated data collection, automated data analysis, and semi-automated data analysis. A wrist activity tracker gathering heart rate, speed, moved distance, and steps per second was proposed for the automated data collection. The basic work plan for extracting direct work hours from the collected data was to identify, understand, and input the specific patterns that appear in each of the different activities in the data analysis process. The patterns were mainly identified and analyzed by a manual process. Then, the patterns were coded in MATLAB for the automatic assignment of activity types to each instance. In addition, J48 decision tree classification in WEKA was used to evaluate how clearly the patterns from collected data have for prediction of activity type.

For demonstration of the proposed method, field experiments were conducted with carpenters at three different construction sites. Results from actual observations (ground truth), J48 classification, and semi-automated data analysis per single experiment were compared to evaluate the proposed method. The comparison showed that the two data analysis methods were able to determine activity types for a laborer wearing a wrist activity tracker. This means that the data collected from the activity tracker has clear patterns that are shown for each activity category type.

5.1. CONTRIBUTIONS

An activity tracker can automatically collect the worker time utilization data without an external observer watching the work activities of a laborer. The data collected can be analyzed at any time while the laborer is working, due to wireless synchronization from the wrist device, a smart phone, and web-based storage. The wrist device will continue to collect data while the synchronized data is analyzed. Therefore, there is no need to designate a time for data collection.

Regarding the quality of input/output, the proposed method can provide considerably objective data with few errors. Once an individual laborer's transition heart rate and patterns of data according to activity types on the job site are firmly established at the beginning, it is doubtful that the productivity rate data could be manipulated. The only way to increase direct work rates would to maintain a high heart rate. This would be possible for a short time, but intentionally maintaining a high heart rate would require that muscles, organs, or nerves be used by the laborer, causing strain and fatigue. Consequently, the data from the proposed method should accurately reflect what the laborer does, providing an objective productivity rate.

The proposed measurement method provides construction field information in as much detail as possible. The wrist device records time, location, and the laborer's condition per second. This provides both advantage and disadvantage to both management and laborers. When something goes wrong during a construction project, management can access detailed information to determine the responsibility of the problem by analyzing the data. From the laborer's perspective, data collection can protect labor by documenting their performance on the job site. With this information, it is possible that some disputable performance problems on construction sites could be resolved objectively with data.

5.2. LIMITATIONS

One of the biggest limitations for individual data collection is the required location of the activity tracker on the body. The wrist is the part of body that is used most during construction activities. The accelerator is very sensitive to rapid swings and whipping motions, so it is important that the activity tracker be worn on the lesser used wrist: left wrist for right-hander and right wrist for left-hander. Additionally, the sensor for measuring heart rates must be attached to the wrist. This means that the wrist band should be a little tight on the wrist, which can cause discomfort.

Other limitations are the number of samples and types of work. This thesis is an exploratory preliminary research study that proposes a new productivity measurement methodology that has never been studied. Thus, the author experimentally conducted the new productivity measurement technique with just a few experiments to determine its feasibility. Only three carpenters joined in the experiments done for this research. More data should be collected to extract firmer standard patterns and validate the methodology. Research is also needed to conduct these proposed methods with different types of construction work. Formwork was the only type of work in the experiments. Therefore, it is possible that the specific patterns detected for activity classification during this research are limited to the formwork activity.

Moreover, in this research, a single person's productivity at a project was calculated and provided. In order to evaluate crew level productivity or project level productivity, many instances of personal productivity of laborers should be collected and integrated.

Lastly, a user interface for data analysis that can be operated by everyone with basic instructions must to be developed. The process of the new approach requires user training to understand and implement the method. The method utilizes the J48 classification in WEKA and MATLAB coding to classify type of works according to pattern. Without basic knowledge of classification and MATLAB language, it is difficult to analyze raw data collected from a wrist activity tracker. Therefore, user-friendly interface that automatically preprocesses raw data, recognizes patterns, classify type of activity, and provide productivity should be developed.

5.3. FUTURE RESEARCH

The new productivity measurement method could be significantly improved with additional research and testing. First, the number of activity categories should be more detailed and refined. The four activity categories that the author defined in this research are too broad as categories to provide the specific information that is needed at construction sites. For example, the “Moving” activity can be subcategorized into “Traveling”, and “Loitering”. The “Moving (High HR)” activity can be specified as “Transporting” and “Handling Material”,

Secondly, the demonstration process should be conducted on various types of work, such as painting, roofing, paving, etc. After additional testing and data collection, the patterns could be standardized according to different types of work. Then, the standardized patterns can be adjusted according to work environment, type of laborer, and type of a construction project when management applies the patterns to their projects.

Finally, the proposed productivity measurement methodology should be implemented on sufficient samples at the crew or project level. A single or a few selections of personnel productivity on a construction project cannot represent the overall performance for the project. After the new productivity methodology become sufficiently mature and reliable, the next step should be to conduct additional research to develop a methodology to calculate a crew’s and project’s productivity from many instances of laborers’ personal productivity.

References

Benzekry, M. (2010). "Identification and Analysis of Practices that Positively Impact Construction Productivity." Master Thesis, The University of Texas at Austin, Austin, TX.

Berndt, E., and Wood, D. (1975). “Technology, Prices and the Derived Demand for Energy.” *Review of Economics and Statistics*, August 1975, pp. 259–268.

Barbero-Alvarez, J. C., Soto, V. M., Barbero-Alvarez, V., and Granda-Vera, J. (2008). “Match analysis and heart rate of futsal players during competition.” *Journal of Sports Sciences,* 26(1), pp. 63-73.

Castellano, J., and Casamichana, D. (2010). “Heart Rate and Motion Analysis by GPS in Beach Soccer.” *Journal of Sports Science and Medicine*, 9, pp. 98-103

Construction Industry Institute (CII). (2006). “Work force view of construction labor productivity.” RR215-11, CII, Austin, TX.

Domar, E. (1961). “On the Measurement of Technological Change,” *Economic Journal*, December 1961, pp. 709–729.

Gatti, U. C., Migliaccio, G.C., and Schneider, S. (2011). “Wearable Physiological Status Monitors for Measuring and Evaluating Worker’s Physical Strain: Preliminary Validation.” *Computing in Civil Engineering, 412(24)* pp. 194-201

Galindo Garre F., De Vries SI. (2012). "Evaluation of Neural Networks to Identify Types of Activity Among Children Using Accelerometers, Global Positioning Systems and Heart Rate Monitors." *An International Perspective on Topics in Sports Medicine and Sports Injury,* ISBN 978-953-51-0005-8, February 2012, pp. 245-256

Gouett, M., Haas, C., Goodrum, P., and Caldas, C. (2011). “Activity Analysis for Direct-Work Rate Improvement in Construction.” *Journal of Construction Engineering and Management*, 137(12), pp. 1117–1124.

Hsiao, W. T., Wu, H. T., and Cheng, T. M. (2012). Construction Research Congress, ASCE, pp. 209-216

Jorgenson, D., and Griliches, Zvi. “The Expansion of Productivity Change,” *Review of Economic Studies*, July 1967, pp. 249–283.

Jorgenson, D., Gollop, F., and Fraumeni, B. (1987). "Productivity and U.S. Economic Growth." *Harvard University Press*, Cambridge, MA.

Kim, J. Y. (2015). "Rapid and Contextual Activity Analysis: A Semi-automated Activity Category, Time, Location, and Video Data Collection and Analysis Methodology." Doctoral dissertation, The University of Texas at Austin, Austin, TX.

Liou, F., and Borcherding, J. (1986). "Work Sampling Can Predict Unit Rate Productivity." *Journal of Construction Engineering and Management*, 112(1), pp. 90–103.

Loftin, M., Anderson, P., Lytton, L., Pittman, P. M., Warren, B. (1996). "Heart Rate Response during Handball Singles Match-play and Selected Physical Fitness Components of Experienced Male Handball Players." *Journal of sports medicine and physical fitness*, 36, pp. 95-99

Oglesby, C. H., Parker, H. W., and Howell, G. A. (1989). "Productivity improvement in construction." McGraw-Hill, New York, NY.

Park, H. S., Thomas, S. R., and Tucker, R. L. (2005). "Benchmarking of construction productivity." *Journal of Construction Engineering and Management*, ASCE, Vol. 131-7, pp. 772-778.

Park, H.-S. (2002). "Development of a construction productivity metrics system (CPMS)."

Maker, R (2016). "Garmin's new Vivoactive HR & Vivofit 3: Hands-on & First Run." DC RAINMAKER, Available at: https://www.dcrainmaker.com/2016/02/garmins-vivoactivehr-vivofit3.html. Last access: Aug. 1, 2017.

Robert Solow, "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, August 1957, pp. 312–320.

See "Multifactor Productivity Trends in Manufacturing, 2006," news release USDL 08-0857 (Bureau of Labor Statistics, May 1, 2008), Available at: www.bls.gov/news.release/archives/prod5_05012008.pdf. Last access: June. 31, 2017.

Sonmez, R., and Rowings, J. E. (1998). "Construction labor productivity modeling with neural networks." *Journal of Construction Engineering and Management*, 124(6), pp. 498–504.

Thomas, H. R., and Holland, M. P. (1980). "Work Sampling: A Comparative Analysis." *Journal of the Construction Engineering and Management Division*, ASCE, 106(4), pp. 519-534.

Thomas, H. R. (1981). "Construction work sampling." Business Roundtable, Oct 1981.

Thomas, H. R., and Daily, J. (1983). "Crew Performance Measurement via Activity Sampling." *Journal of the Construction Engineering and Management Division*, ASCE, 109(3), pp. 309-320

Thomas, H. R., and Mathews, C. T. (1986). “An analysis of the methods for measuring construction productivity.” Source document No. 13, Construction Industry Institute (CII), Austin, TX.

Thomas, H. R., Maloney, W. F., Horner, R. M. W., Smith, G. R., Handa, V. K., and Sanders, S. R. (1990). “Modeling Construction Labor Productivity.” *Journal of Construction Engineering and Management,* 116(4), pp. 705–726.